

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2002, 10:39:25 ; Search time 37.13 Seconds

(without alignments)
1932.499 Million cell updates/sec

Title: US-09-653-961-2

Perfect score: 3363
Sequence: 1 MGLPRLVCATLAAACCCPR.....SSGDKRAPDGGKTYIDLRH 646

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802:*
1: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID55/gcgcdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID55/gcgcdata/geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3341	99.3	646	20	AAV42404
2	3327	98.9	646	20	AAV42405
3	3327	98.9	646	22	AAAB90798
4	647.5	19.3	628	22	ABB50259
5	647.5	19.3	628	22	ABB50290
6	602	18.3	136	21	AAAG00308
7	602	17.9	595	22	ABR11999
8	498.5	14.8	574	21	AAV93969
9	494.5	14.7	582	22	AAAB76858
10	494.5	14.7	583	17	AAAB7230
11	492.5	14.6	583	17	AAW06891

12	492.5	14.6	583	19	AAW47088	Activated leukocyte
13	466	13.9	570	17	AAE97231	Stem cell marker H
14	414.5	12.3	394	21	AAAB3963	Human cancer assoc
15	280.5	8.3	749	22	ABR10297	Human cDNA seq ID
16	277.5	8.3	712	22	AAE07070	Human gene 20 enco
17	269.5	8.0	550	22	ABR10486	Human cDNA seq ID
18	269.5	8.0	550	22	AAU17986	Human immunoglobul
19	266	7.9	3117	22	AAV53667	Sequence g1/332818
20	264.5	7.9	534	21	AAV95323	Pig VCAM, Sus scr
21	253	7.5	269	22	ABR10346	Human cDNA seq ID
22	252	7.5	1380	20	AAV08402	Drosophila sp. ROB
23	252	7.5	1381	20	AAV13564	Drosophila Robo 2
24	246.5	7.3	600	22	AAU29316	Human PRO polypept
25	244.5	7.3	647	14	AAV61552	VCAM-6D/ICAM-1.
26	244.5	7.3	823	22	ABR61552	Drosophila melanog
27	237.5	7.1	736	14	AAV39682	VCAM/ICAM-2. Synt
28	236	7.0	1240	22	ABG03933	Novel human diagno
29	236	7.0	4412	21	AAV53666	Sequence g1/101742
30	235.5	7.0	4393	22	AAV31889	Amino acid sequenc
31	235.5	7.0	4436	22	ABG23285	Novel human diagno
32	234	7.0	848	21	AAV88565	Human NCAM 140KD
33	233	6.9	647	11	AAV08117	Human p53 target m
34	233	6.9	647	14	AAV38549	Human p85 protein.
35	233	6.9	647	21	AAV59498	Rabbit ATRERO-ELAM
36	233	6.9	1496	20	AAV81030	Drosophila melanog
37	233	6.9	1496	21	AAV70469	Melanoma associate
38	233	6.9	1498	22	ABR11587	Human p53 target m
39	231.5	6.9	753	20	AAV63927	Human p85 protein.
40	231.5	6.9	828	12	AAV13905	Human 185 protein.
41	230.5	6.9	968	22	ABR60068	Rabbit ATRERO-ELAM
42	230	6.8	735	14	AAV38550	Drosophila melanog
43	230	6.8	739	11	AAV08118	VCAM/ICAM-1. Synt
44	230	6.8	739	12	AAV10316	Vascular cell adhe
45	230	6.8	739	13	AAV21081	1E7/2G7 stiaologlyc

ALIGNMENTS

RESULT	1	
AAV42404		
ID	AAV42404 standard; Protein; 646 AA.	
XX		
AC	AAV42404;	
XX		
DT	01-DEC-1999 (first entry)	
XX		
DE	Human MUC18 amino acid sequence.	
XX		
KW	prostate cancer; melanoma; cell adhesion; glycoprotein;	
KW	metastasis; treatment; detection; diagnostic test.	
OS	Homo sapiens.	
XX		
PN	WO9945392-A1.	
XX		
PD	10-SEP-1999.	
XX		
PF	02-MAR-1999; 99WO-US04850.	
XX		
PR	03-MAR-1998; 98US-0076664.	
XX		
PA	(UYEM-) UNIV EMORY.	
XX		
PI	WU G;	
XX		
DR	WPI, 1999-540899/45.	
XX		
PT	N-PSDB; AA20930.	
XX		
PS	Detection of metastatic prostate cancer, by detection of MUC18 expression in prostate cancer cells -	
PS	Claim 17; Page 7; 80pp; English.	

XX This is the amino acid sequence of the Human MUC18 cell adhesion
CC glycoprotein, which is expressed on the surface of melanoma cells,
CC and can be used as a marker for prostate cancer.
CC This MUC18 (hMUC18) cDNA sequence is different from the hMUC18 cDNA
CC given in Genbank accession number AAN28882, AAZ20931. The deduced amino
CC acid residues from this cDNA are very different from that published by
CC Johnson's group in 1989.
CC The presence of this glycoprotein has been correlated with the ability
CC of melanomas to metastasize. MUC18 is also associated with normal
CC vascular tissue, and on the smooth muscle of venules, and it expresses
CC sporadically on capillary epithelium.
CC The method can be used as a diagnostic test for prostate cancer which
CC has a relatively high potential for metastasis or which has
CC metastasized. The physician can then choose the appropriate surgical,
CC chemotherapeutic or radiation treatment regimens. In addition, antibody
CC specific to MUC18 can be used to prevent metastasis of PCs.

XX
SQ Sequence 646 AA;

Query Match 99.3%; Score 3341; DB 20; Length 646;
Best Local Similarity 99.5%; Pred. No. 6.8e-227;
Matches 643; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MGLPRIVCAFLLAACCCPRVAGVGEAOPAPELVEVEVSTALKGSLQSOGNLSHV 60
DB 1 mglprlvcafllaaaccprvagvpgeapelpelveevgstallkqslsgqgnlshv 60
QY 61 DMFSVHKERTLLIFRVROGOGSEPEYEDORSLODRGATLALTOVTPQDERIFLCQGR 120
DB 61 dmfsvhkerllifrvrvgogsepeyedrslodrgatlaltovtpqderiflfcqgr 120
QY 121 PRSOEYRIOLRYVKAPEEPNIOVNPGLIPVNSKEPEEVATCGRNQYPIPOVWYKNGRP 180
DB 121 prsgeyriqlryvkapeepniqvnplgipvnskepeevatcgrngypipqvwykngrp 180
QY 181 LKEEKNRVHIQSSQTVESGLTYLTQSTLKAQLYKEDKDAOFYCELYNRLPSGNHMKESRE 240
DB 181 lkeeknrhvhiqssqtvesgltlyltqstlkaqlvkedkdaqfycelynrlpsgnhmkesre 240
QY 241 VYVPVFPRTKWLVEVPVWMLKEGDRVETRCIADNPPRHFSISQNSQSTRAREETIN 300
DB 241 vyvpvfprtkwlvewpvmllkegdrvetrciadsnprrhfsisqnsqstrareeetn 300
QY 301 DNGVLVEPARKHSGRVECOGLDLDTMISLSEPOELLVNYSVDVRSFAPAREQEGSS 360
DB 301 dngvllveparkhsgrveecogldldtmislsepegellvnysvdvrspareqegss 360
QY 361 LTTTCEAESSQDLEFQWLREETSQVLERGPVLOLHDLKREAGGGRVAVSPSIPGLNRT 420
DB 361 ltttceassqdlefqwlreetqvlergpvlqlhdklreagggrvavspspipglnrt 420
QY 421 QLVNVAIFGPPMAFKRKRWVKNWVNLSCFASGHPRTTSMNNGTSEDODOPOR 480
DB 421 qlvnvaifgppmafkrkvwknwvnlscfashprrttismnngtseodoporr 480
QY 481 LSTLNVLVTPPELLETGVECTASNDLGKNTSILFLLELVNLTLLTPDSNTTGLSTASPH 540
DB 481 lstlnvlvtppeletgvectasndlgkntsilfllelvnltlltpdsnttglstasph 540
QY 541 TRANSTSTERKLPPEESRGVYIAVYICILVAVLAVKLVFLKKGKLPFRSGKQETL 600
DB 541 tranststerklppeesrgvyiavyicilvavlvavlkvflkkgklprrisgkqetl 600
QY 601 PPSKSELVVEVSKDKLPEEMGLLOGSSGDKRAPGDGEGYIDLRH 646
DB 601 ppskselvvvskdklpeemgllqsssgdkrapgdgqekyidlrh 646

RESULT 2
AAV42405
ID AAV42405 standard; Protein; 646 AA.

XX AAV42405;
AC 01-DEC-1999 (first entry)
DT Human MUC18 amino acid sequence, (Johnson 1989).
DE prostate cancer; melanoma; cell adhesion; glycoprotein;
KW metastasis; treatment; detection; diagnostic test.
OS Homo sapiens.
XX MO9945392-A1.
XX 10-SEP-1999.
XX 02-MAR-1999; 99MO-US04850.
XX 03-MAR-1998; 98US-0076664.
XX (UYEM-) UNIV EMORY.
XX Wu G;
XX WPI: 1999-540899/45.
XX N-PSDB: AAZ20931.
XX Detection of metastatic prostate cancer, by detection of MUC18
XX expression in prostate cancer cells -
XX Disclosure: Page 8; 80pp; English.
XX This is the amino acid sequence of the Human MUC18 cell adhesion
XX glycoprotein, which is expressed on the surface of melanoma cells,
XX and can be used as a marker for prostate cancer.
XX This sequence was originally published in 1989 by Johnson et al
XX (Genbank AAN28882).
XX This sequence differs from the one proposed by the inventors, this amino
XX acid sequence is 603 amino acid residues while the present sequence
XX AAZ20930 has 646 amino acid residues.
XX The presence of this glycoprotein has been correlated with the ability
XX of melanomas to metastasize. MUC18 is also associated with normal
XX vascular tissue, and on the smooth muscle of venules, and it expresses
XX sporadically on capillary epithelium.
XX The method can be used as a diagnostic test for prostate cancer which
XX has a relatively high potential for metastasis or which has
XX metastasized. The physician can then choose the appropriate surgical,
XX chemotherapeutic or radiation treatment regimens. In addition, antibody
XX specific to MUC18 can be used to prevent metastasis of PCs.

SQ Sequence 646 AA;

Query Match 98.9%; Score 3327; DB 20; Length 646;
Best Local Similarity 98.9%; Pred. No. 6.5e-226;
Matches 639; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MGLPRIVCAFLLAACCCPRVAGVGEAOPAPELVEVEVSTALKGSLQSOGNLSHV 60
DB 1 mglprlvcafllaaaccprvagvpgeapelpelveevgstallkqslsgqgnlshv 60
QY 61 DMFSVHKERTLLIFRVROGOGSEPEYEDORSLODRGATLALTOVTPQDERIFLCQGR 120
DB 61 dmfsvhkerllifrvrvgogsepeyedrslodrgatlaltovtpqderiflfcqgr 120
QY 121 PRSOEYRIOLRYVKAPEEPNIOVNPGLIPVNSKEPEEVATCGRNQYPIPOVWYKNGRP 180
DB 121 prsgeyriqlryvkapeepniqvnplgipvnskepeevatcgrngypipqvwykngrp 180
QY 181 LKEEKNRVHIQSSQTVESGLTYLTQSTLKAQLYKEDKDAOFYCELYNRLPSGNHMKESRE 240
DB 181 lkeeknrhvhiqssqtvesgltlyltqstlkaqlvkedkdaqfycelynrlpsgnhmkesre 240

QY 241 VTVPVFTPEKVMLEVEPVGMLEKGDRIEIRCLADGNPPHFSISKONPSTREAEEETT 300
 Db 241 vtvpvftpekwleevpvmlekedriveirciadgnpphfsisknpstreaeeetn 300
 QY 301 DNGVLVLEPARKHSGRTECGLDLDTMTLSLSEPOELLVNVYSDVRVSPAAPEROGSS 360
 Db 301 dngvlvleparkhsgryecqawldtmtslisepellvnyvsdvrvspaaperogss 360
 QY 361 LTLTCEAESSODLEFOWMLREETGOVLERGVLQDLHLKREAGGGRCAVSPISPGINRT 420
 Db 361 ltlceaeessodlefgwmlreetdqvlergvqlndhlkreagggyrcvaspsipglnrt 420
 QY 421 QLVNVAIFGPPWMAFKERKVVWKENMVNLSCASGHPRTISWVNGTASSEDODOPORV 480
 Db 421 qlvnlafgppwmafkervvkenmvnlscasghprtlswvngtastsedodopqr 480
 QY 481 LSTLNVLTPPELLETGVECTASNDLGKNTSIFLELVNLTTLTPSNTTGTASTASPH 540
 Db 481 lstlnvltpelleltgvectasndlgkntsifllelvnlttltpsdnttgtastasp 540
 QY 541 TRANSTSTERKLPEPESRGVIVAVICILVAVLGAVALYFLYKRGKLPCCRSGKQETTL 600
 Db 541 tranststerklpepesrgvivavivciltvavlgavlyflykrgklpcrrsgkqetl 600
 QY 601 PPSRKSELVVEVKSDDLPEEMGLIOGSSGDKRAPDQGEKTYDLRH 646
 Db 601 ppsrkcelvvevksdklpeemgllqssgdkrapdqqektydlrh 646
 RESULT 3
 AAB90798
 ID AAB90798 standard; Protein; 646 AA.
 AC AAB90798;
 DT 15-JUN-2001 (first entry)
 DE Human shear stress-response protein SEQ ID NO: 96.
 KW Human; shear stress-response protein; vascular disease;
 KW arteriosclerosis.
 OS Homo sapiens.
 PN WO200125427-A1.
 PD 12-APR-2001.
 PE 02-OCT-2000; 2000WO-JP06840.
 PR 01-OCT-1999; 99JP-0280976.
 PA (KYOM) KYOMA HAKKO KOGYO KK.
 PA (NOJIR) NOJIMA H.
 PI Nojima H, Yoshisue H, Obayashi M, Ota T, Kawabata A, Sakurada K;
 PI Kuga T, Sekine S, Nakamura Y, Sugano S;
 DR WPI; 2001-266308/27.
 DR N-PSDB; AAH02921.
 PT DNA sequences, proteins encoded by them and antibodies against them
 PT useful in diagnosis and treatment of vascular disease caused by
 PT arteriosclerosis -
 XX Claim 60; Page 497-500; 678pp; Japanese.
 CC The present invention provides the protein and coding sequences of a
 CC number of human shear stress response proteins. These are useful in the
 CC diagnosis, treatment and screening of vascular diseases caused by
 CC arteriosclerosis, including heart failure, post-PTCA restenosis and
 CC hypertension.
 XX

SQ Sequence 646 AA.
 Query Match 98.9%; Score 3327; DB 22; Length 646;
 Best Local Similarity 98.9%; Pred. No. 6.5e-226;
 Matches 639; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MGLPRLVCAFLLAACCCCEPRVAVGPDEAPAPAEIYVEVGTALTKGSLQSOGNLSHV 60
 Db 1 mglprlvcafliaaccccpvavgpdeapapaeivevgtallkcglsqsgnls 60
 QY 61 DMFSYHKERTLIFRRYOGOGSSEPEYRRLSLQDRGATLALQTVPDERTFLCGGR 120
 Db 61 dmfsyhkerltlifrrogogsspeeyrrslslqdrगतलalqtvpertrflcggr 120
 QY 121 PRSGEYRIQRYKKAPEEBNIOVNPGLGIVNSKEPEEVAFCVRNGPFIQVIMYKNGRP 180
 Db 121 prsgeyriqlryykaapebnigvnpplgivnskepeeatcvgngypilqvilykngrip 180
 QY 181 LKEENRHYIOSQTVESGLYTLQSLKAQLYKEKDKAOFYCELNRYLPSGNHMKESRE 240
 Db 181 lkeeknrhyiosqtvesgyltqlslkaqlvkekdkaofycelnrylpsgnhmkesre 240
 QY 241 VTVPVFTPEKVMLEVEPVGMLEKGDRIEIRCLADGNPPHFSISKONPSTREAEEETT 300
 Db 241 vtvpvftpekwleevpvmlekedriveirciadgnpphfsisknpstreaeeetn 300
 QY 301 DNGVLVLEPARKHSGRTECGLDLDTMTLSLSEPOELLVNVYSDVRVSPAAPEROGSS 360
 Db 301 dngvlvleparkhsgryecqawldtmtslisepellvnyvsdvrvspaaperogss 360
 QY 361 LTLTCEAESSODLEFOWMLREETGOVLERGVLQDLHLKREAGGGRCAVSPISPGINRT 420
 Db 361 ltlceaeessodlefgwmlreetdqvlergvqlndhlkreagggyrcvaspsipglnrt 420
 QY 421 QLVNVAIFGPPWMAFKERKVVWKENMVNLSCASGHPRTISWVNGTASSEDODOPORV 480
 Db 421 qlvnlafgppwmafkervvkenmvnlscasghprtlswvngtastsedodopqr 480
 QY 481 LSTLNVLTPPELLETGVECTASNDLGKNTSIFLELVNLTTLTPSNTTGTASTASPH 540
 Db 481 lstlnvltpelleltgvectasndlgkntsifllelvnlttltpsdnttgtastasp 540
 QY 541 TRANSTSTERKLPEPESRGVIVAVICILVAVLGAVALYFLYKRGKLPCCRSGKQETTL 600
 Db 541 tranststerklpepesrgvivavivciltvavlgavlyflykrgklpcrrsgkqetl 600
 QY 601 PPSRKSELVVEVKSDDLPEEMGLIOGSSGDKRAPDQGEKTYDLRH 646
 Db 601 ppsrkcelvvevksdklpeemgllqssgdkrapdqqektydlrh 646
 RESULT 4
 ABB50259
 ID ABB50259 standard; Protein; 628 AA.
 AC ABB50259;
 DT 08-FEB-2002 (first entry)
 DE Lutheran blood group protein (BCAM) ovarian tumour marker protein, #6.
 KW Ovarian tumour marker gene; human; overexpression; upregulation;
 KW epithelial tumour; cancer; diagnosis; prognosis; disease monitoring;
 KW identification; serous cystadenoma; borderline serous tumour;
 KW serous cystadenocarcinoma; mucinous cystadenocarcinoma;
 KW mucinous cystadenoma; borderline mucinous tumour; endometrioid carcinoma;
 KW undifferentiated carcinoma; clear cell adenocarcinoma; cystadenofibroma;
 KW adenofibroma; Brenner tumour; serial analysis of gene expression; SAGE;
 KW immune response pathway; cell proliferation regulation; protein folding;
 KW membrane localised; secreted; therapeutic target; cytostatic;
 KW gene therapy; vaccine.
 XX

OS Homo sapiens.
 XX MO200175177-A2.
 XX 11-OCT-2001.
 PD 03-APR-2001; 2001WO-US10947.
 XX 03-APR-2000; 2000US-194336P.
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Morin PJ, Sherman-Baust CA, Pizer ES, Hough CD;
 PI WPI; 2001-626450/72.
 DR N-PSDB; ABA83083.
 XX
 PT Detecting and identifying ovarian tumor, identifying increased risk for
 PT developing ovarian cancer, and determining effectiveness of ovarian
 PT cancer treatment, by measuring expression level of ovarian tumor marker
 PT gene
 XX
 PS Claim 22: Page 55-56; 140pp; English.
 XX
 CC The invention relates to methods for diagnosing and prognosing ovarian
 CC tumours in an individual via the detection and measurement of the
 CC expression of ovarian tumor marker genes (ABA83081-ABA83122, ABA83180,
 CC ABA83182 and ABA83184) or segments thereof (ABA83123-ABA83169, ABA83179,
 CC ABA83181 and ABA83183). The methods of the invention are useful for
 CC detecting an ovarian tumor in a patient, for identifying an individual
 CC at increased risk for developing ovarian cancer, in prognostic tests for
 CC assessing the relative severity of ovarian cancer, in tests for
 CC monitoring a patient in remission from ovarian cancer and in tests for
 CC monitoring disease status in a patient being treated for ovarian cancer.
 CC The methods can additionally be used to identify a particular tumour as
 CC being an ovarian tumour (i.e., an epithelial ovarian tumour selected from
 CC serous cystadenoma, borderline serous tumour, serous cystadenocarcinoma,
 CC mucinous cystadenoma, borderline mucinous tumour, mucinous
 CC cystadenocarcinoma, endometrioid carcinoma, undifferentiated carcinoma,
 CC clear cell adenocarcinoma, cystadenofibroma, adenofibroma and Brenner
 CC tumour. The ovarian tumour marker genes of the invention were identified
 CC using SAGE (serial analysis of gene expression) and were found to be
 CC overexpressed in a broad variety of ovarian epithelial tumour cells
 CC relative to normal ovarian epithelial cells. The marker genes are
 CC implicated in immune response pathways, in the regulation of cell
 CC proliferation and in protein folding, and many of these are membrane-
 CC localised or secreted. In addition to their use as diagnostic and
 CC prognostic markers, the ovarian tumour marker genes or their encoded
 CC proteins may be used as therapeutic targets for the treatment and
 CC prevention of ovarian cancer. Sequences ABB50257-ABB50299 represent
 CC proteins encoded by ovarian tumour marker genes of the invention.
 XX
 SO Sequence 628 AA;

Query Match 19.3%; Score 647.5; DB 22; Length 628;
 Best Local Similarity 29.5%; Pred. No. 3.6e-37;
 Matches 202; Conservative 103; Mismatches 265; Indels 115; Gaps 27;

QY 2 GUPR-LVCAFLAACCCCRVAGVPGAEQAPABELVEVEVSGSTALKCGISQSGNLSH- 59
 DB 12 GPRILLIALLAVIAA-----hpdagaevrlsvpplvevmgkxvlldc---tpgtctdh 62
 QY 60 -VDWF-----SVHKEKRTLIFRVROGOGSEPEYBORSLDRGATLALITQV 106
 DB 63 mlewltldsgarprlasaemgseqlvmbdtrgrsp-----yqldsgqr---lvlaaa 115
 QY 107 TPQDERIFLC--QGKRRPOEYRIQLRVYKAPPEPNIQVPLGIPVNSKPEEVAACGR 164
 DB 116 gvgderdyvcvragaaataarlnvtfakpeatevspkkytcltsvmedsaqelactstr 175
 QY 165 NGYPIQVYIMYKNGRPLK--EKKNRVHIOSSQTV-ESSGLYTLQSLIKALQVKKDKDAOF 221

DB 176 ngnpapklitwyrnggrllevvempdegymtstrtvreaqllslstcljrlrkddasf 235
 QY 222 YCELNRYLPSCGHHMK-ESREYVVPVFPYREKY--WL--EVEPYGMKEDGRVFRICADG 276
 DB 236 hcaahyslpegahridpsltfhlthylehvfhwgspstpegwvregdtvqllcrqdg 295
 QY 277 NPPHFISKQNPSTREAEETNDN--GVILEPARKBHSGRYECOGDLDOTMISL-IS 333
 DB 296 spspeytlfr-----lqdegevlvnalegnltlegvtirgsgfygcrvedydaadvqls 351
 QY 334 EPOELLVNVSVSVRVPAPAEQEG-----SSLTICEAASSODLERQWLREFTGVY 365
 DB 352 kclrlvayldp|els-----egkvlslplnssavvnosvnhjlpalrlwtcdst--p 402
 QY 386 LERGPVLOLHDKREAGGRCVAVSPSIPGLNRTOLVNAVIRPMPMAFERK-----V 440
 DB 403 lqdgpmllssltfdngcyceaslprrvlarstqntfllvgspklcaalepkeds 462
 QY 441 WYKENVNLNISEASGHPRTISWVNGTASQEDQDPQ--VLSTINLVTPLELTGV 497
 DB 463 w-regdevcllcsarghpbdklswsqly-gspaeplpgrgwwsssltkvtsalsrdyl 520
 QY 498 ECTASNDLGNKNSILFLELVNLTLPDSTTTGLSTASPHTRANSTSTERKLPPES 537
 DB 521 sceasnphnknkhvthfgavs-----pqtsq 546
 QY 558 RGVIVAVIYICILVAVLGLVLYLYKRGKLPORRSKQETLPPSRKSELVEYKSDKL 617
 DB 547 agvaavnavavsvgl|lllvavvfycvrrkkg-pccrgrrekgapp---gepglshsgsq 602
 QY 618 PEMEGLLQG--SSGDKRAPDQGEK 640
 DB 603 pegtglmgsagsgarqysgfigde 627

RESULT 5
 ABB50290 standard; protein; 628 AA.
 ID ABB50290;
 XX ABB50290;
 AC
 XX
 DT 08-FEB-2002 (first entry)
 XX
 DE Lutheran blood group (BCAM) ovarian tumour marker protein, SEQ ID NO:70.
 XX
 KW Ovarian tumour marker gene; human; overexpression; upregulation;
 KW epithelial tumour; cancer; diagnosis; prognosis; disease monitoring;
 KW identification; serous cystadenoma; borderline serous tumour;
 KW serous cystadenocarcinoma; mucinous cystadenocarcinoma;
 KW mucinous cystadenoma; borderline mucinous tumour; endometrioid carcinoma;
 KW undifferentiated carcinoma; clear cell adenocarcinoma; cystadenofibroma;
 KW adenofibroma; Brenner tumour; serial analysis of gene expression; SAGE;
 KW immune response pathway; cell proliferation regulation; protein folding;
 KW membrane localised; secreted; therapeutic target; cytostatic;
 KW gene therapy; vaccine.
 KW
 XX
 OS Homo sapiens.
 XX
 XX MO200175177-A2.
 XX 11-OCT-2001.
 PD 03-APR-2001; 2001WO-US10947.
 XX 03-APR-2000; 2000US-194336P.
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Morin PJ, Sherman-Baust CA, Pizer ES, Hough CD;
 PI WPI; 2001-626450/72.
 DR N-PSDB; ABA83116.

Db 528 sceasn-----phgnkrhvfhfgtvsqpt-----sq 553

QY 558 RGVTAVAVICLVAVLAVGAVLYFLTKKGKLPGRSGKOE 597

Db 554 agvavmavavsvglillvvvfyvcrkrkg-pccqrgrrek 592

RESULT 8

AA93969

ID AA93969 standard; Protein: 574 AA.

AC AA93969;

DT 03-OCT-2000 (first entry)

DE Amino acid sequence of a partial rat antigen Ag2117.

XX Antigen: Ag2117; pancreatic cell; monoclonal antibody; surface antigen.

XX Rattus sp.

OS WO200037503-A1.

PN 29-JUN-2000.

PD 22-DEC-1999; 99WO-US30741.

PF 22-DEC-1998; 98US-0218539.

PR 22-DEC-1998; 98US-0218539.

XX (RAVE-) RAVEN BIOTECHNOLOGIES INC.

PA Mather JP, Bald LN, Roberts PR, Stephan JF;

PI MPI: 2000-452179/39.

XX N-PSDB; AA57480.

DR Immunizing a host mammal to produce population of monoclonal antibodies

XX that bind to antigens of specific cell type comprises introducing

PT viable cells with serum-free surfaces -

XX

PS Example 10; Fig 10; 61pp; English.

XX The present sequence represents rat antigen Ag2117. The antigen is

CC representative of pancreatic cell lines. The antibody that recognises

CC Ag2117 was identified using the method of the invention. The

CC specification describes a method for immunizing a host mammal to produce

CC a population of monoclonal antibodies that bind to antigens

CC representative of a specific cell type that are heterologous to the

CC host mammal. The method comprises introducing into the mammal intact

CC and viable cells with surfaces which are free of serum. The methods

CC are useful for producing populations of antibodies that bind to antigens

CC representative of a specific cell type, specifically antibodies that

CC are tissue-selective, sub-tissue selective or cell-type specific. These

CC antibodies facilitate the identification of novel antigens and the

CC delineation of the combination of surface antigens on a specific cells

CC type. The antibodies or hybridomas produced using the methods have

CC diagnostic and therapeutic applications.

CC

SO Sequence 574 AA:

Query Match 14.8%; Score 498.5; DB 21; Length 574;

Best Local Similarity 24.7%; Pred. No. 1e-26;

Matches 151; Conservative 122; Mismatches 265; Indels 73; Gaps 25;

QY 5 RVNAFLAACCCCPRAVPGAEAPAPLVEVVGSTLLKCGLSQSGNLSHWDWS 64

Db 10 rlvfccllisaavlrpgjgwy-----tvsaygdltvmpcorldvpq-nlmfgrkwy 58

QY 65 VHKERTLLIFRRVQGGQS-----EPGEYEQRLSDRGATLALVOYTPQDERIFLQ-QK 119

Db 59 ekpdpvflafirstctktskvgyddvpeykarlrlse-nylsltnaklsdektfvomlv 117

QY 120 RPRSQEYRIOLRYVAKAPEEENIOWNPGLIPVNSKEBEENATCGNRGPIPOVIWYKNGR 179

Db 118 ednufaeaprlvkrfkqskpeil-vnraaf-leteqklkljgslrdsydgntwyrngk 175

QY 180 PLKEERKRVHIIQSSQIYE-SSGLITLOSILKADLVEDKDAQYCYCLNRLPSGNIMKES 238

Db 176 vlgpvdgevalllfkkeidpqtqlytmssleykctksdiqmpitcsvtyygsqgktyis 235

QY 239 REVTVPVEYFPEKRWLEEV-EPVGMLEKGDREVEIRCLADGNPPH---FSLKONPSTREA 294

Db 236 eqaifdiyyregevtivlppknaikgedniltqclgnqnppeetlmfjlpqgaegirs 295

QY 295 EEETNDNGVLYLEPARKESGRYECOGDLDITMISLSEPOELVNVYSDVAVSPAP- 353

Db 296 ntyltd-----vratnagdkcalldqrmaast---ltvhyldslnpsgev 342

QY 354 ERQGSSTLTCEASSODLEFQWLREPEQVLEERGVLQDLHKKREAGGCGYCAVASVS 413

Db 343 tkqldtclpvscltsaarnatvwmkdnlr-irssp-sfsalhyqdaqnyccetaige 398

QY 414 IPGINRTQLVNVALFGPPMAFAKERKVVKENVNLNLSCEASGHPRTTSMVNGTAS-- 471

Db 399 veglkkresltlvegkr-qikmtkktpsglskcltichvegfppkaiqwtlgsav 456

QY 472 -EQDQDPQ--RVLSTNLVTPPELLETGVBCIASNDLGKNTSLILELVNLTTLTPDSN 527

Db 457 nteespyingrys--klispeenvt-ltctaenqlertvsnlsaisl----pe-- 507

QY 528 TTGTGISTASPHRANSTSPERKLPEPSRGVIVAVICIIVLAVGAVFLY-KKG 586

Db 508 -----headdisident-ekvndqakllygvlllaalvagvvywlymkks 554

QY 587 KLPGRSGKOE 597

Db 555 ktaeshakkk 565

RESULT 9

AA976858

ID AAB76858 standard; Protein: 582 AA.

XX AAB76858;

XX 12-APR-2001 (first entry)

DE Human lung tumour protein related protein sequence SEQ ID NO:334.

XX Human: lung cancer; lung tumour; lung tumour protein; gene therapy;

XX lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;

XX cytosolic; antisense inhibition.

XX Homo sapiens.

XX WO200100828-A2.

XX 04-JAN-2001.

XX 30-JUN-2000; 2000WO-US18061.

XX 30-JUN-1999; 99US-0346492.

XX 15-OCT-1999; 99US-0419356.

XX 17-DEC-1999; 99US-0468867.

XX 30-DEC-1999; 99US-0476300.

XX 06-MAR-2000; 2000US-0519642.

XX 22-MAR-2000; 2000US-0533077.

XX 10-APR-2000; 2000US-0546259.

XX 27-APR-2000; 2000US-0560406.

XX 05-JUN-2000; 2000US-0589184.

XX (CORI-) CORIXA CORP.

XX Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;

PI Retter MW, Mannion J;

Db	212	adlqmpfcsvtyyyppsgqkthsegavfdlypbeqvltqvlpbkhaiknegdnitkcl	271
Oy	274	ADGNPPH---FSISKQNPSTREAEETNDNGVLVLEPAKREHSGRYECQGLDPTMIS	330
Db	272	gnunpppeefllylpqperlrsntytltd-----vyrnatgqkcslldkkemia	323
Oy	331	LISEFOELLNVNSDVRSPAP--EROGSSLTLTCEANESODLEFQWLREETGOYLERC	389
Db	324	sta-----ltvlyl dlsinspgevtrtqgdalpvscissrnaatvymmdnir--lrs	376
Oy	330	PVLQHLDKRPAAGGRCVAPSIPGLNRTQLVNVAIFGPPMAAFERKRVWYKENVVLN	449
Db	377	p--sfsslhgdagqnyccetalgveglkkresitllivegkp--qlkmtktdpsjst	432
Oy	450	LSCASNGRPRTIMNNNGTAS---EDDQDPQ---RVLSTLVNLYPPELLETGVECTASN	503
Db	433	ichveglpdkpalwtltqsgvlnqleespylngrtys--kllispeenvt-ltctaen	489
Oy	504	DLGNKTSILFELNLTITLPDSTNTTGLSTASPTANSTRKLPEDPSRGVIT	563
Db	490	qlertvsnlvsalsi-----pe-----hdeadeisdenr-ekvndqakily	530
Oy	564	AVYCIILVAVLAVLYFLY-KKCKLPDRSGSGQETILPSPRSSE	607
Db	531	givvglllaalvagvyvlymkkkskaskhvnkdlgmeenkklle	575
RESULT 11			
ID	AAW06891	standard; Protein; 583 AA.	
XX	AAW06891;		
XX	18-MAR-1997	(first entry)	
DE	Human activated leukocyte-cell adhesion molecule ALCAM.		
XX	Activated leukocyte-cell adhesion molecule; ALCAM; CD6 ligand;		
KW	thymic epithelial cell; inflammation; allograft rejection;		
KW	neurodegenerative disease.		
XX	Homo sapiens.		
OS			
XX			
FH	Key	Location/Qualifiers	
FT	Peptide	1..27	
FT		/label= Sig-peptide	
FT	Protein	28..583	
FT		/label= Mat-protein	
FT	Domain	28..527	
FT		/label= Extracellular_domain	
FT	Domain	/note= "the extracellular domain can be subdivid"	
FT		528..551	
FT	Domain	/label= Transmembrane_domain	
FT		552..583	
FT	Modified-site	/label= Extracellular_domain	
FT		91	
FT	Modified-site	/label= Glycosylation	
FT		/note= "putative N-glycosylation site"	
FT	Modified-site	/label= Glycosylation	
FT		/note= "putative N-glycosylation site"	
FT	Modified-site	/label= Glycosylation	
FT		/note= "putative N-glycosylation site"	
FT	Misc-difference	258	
FT		/label= Asn, Ser	
FT	Modified-site	/note= "variation owing to polymorphism in HL60	
FT		and T-cell derived clones"	
FT	Modified-site	265	
FT		/label= Glycosylation	
FT		/note= "putative N-glycosylation site"	
FT	Misc-difference	301	

[illegible]

```

Db      152 kllgdcisedspdgnticwyrngkvlhpllegavvliifkempdvrcqlytmstleykctk 211
Oy      215 EKDQOFCELNRLPSGNHMKESREVTVPVFPPEKWLVEY-EPVGNLKGDRVEIRCL 273
Db      212 adlqmpftcsvlygypsgqkthseqavfdiypqlegvliqvlppkxakagddnlclxcl 271
Oy      274 ADGNPPPH---PSISKONPSTRAEETNDNGVLTLPARKENHGRECOGLDIDTMS 330
Db      272 gngnpppeefilylpgqpegitssntlytlxd-----vrrnatgdykcsalidcksmia 323
Oy      331 LUSEPOELLVNVSDVRSVPAAP-ERQEGSSITLTCEASSODLEFQWLREBTQVLERG 389
Db      324 sta-----ltvhy1-dlslnpsgevrtrlgdaipvscitssnatvwmkdnir--lrs 376
Oy      390 PVLQHLDLKREAGGRCVASTPSTPGLNRLQVNVAFEGPPWMAFKRKRWVKNEMVLN 449
Db      377 p--sfslhygdagayvcetalaivegllkrteslclivegkp--qikmtktqdsqiskt 432
Oy      450 LSCASGHPRTSMNVNGTAS---EODDPPQ---RVLSTLVNLTVPLELTGVECTASN 503
Db      433 l1chvegfkpalqwtlgsavlnqteespylndryys--klisspeenvt-lcttaen 489
Oy      504 DLGKNTSIFLELVNLTITLPDSNNTTGLSTASTASPHRANSTERRKLPESRGVYIV 563
Db      490 qlertvnslnvasaisl---pe-----hedeaisdent-ekvndqaklly 530
Oy      564 AVIVCIVLAVLGAVALFLY-KKGLPCRRSGKOETITPPSRKSE 607
Db      531 glvvgl1laalvagvywlymkksktaakhvnxkldgmneenkle 575

RESULT 12
AAW47088
ID      AAW47088 standard; Protein; 583 AA.
XX      AAW47088;
AC
XX      20-JUL-1998 (first entry)
DE      Activated leukocyte-cell adhesion molecule (ALCAM).
XX      ALCAM; activated leukocyte-cell adhesion molecule; CD6 ligand;
KW      antibody; human.
XX      Homo sapiens.
OS
XX      Key Location/Qualifiers
FH      1..27
FT      /label= Sig-peptide
FT      28..583
FT      /label= Mat_protein
FT      28..527
FT      /note= "extracellular domain"
FT      528..551
FT      /note= "transmembrane domain"
FT      552..583
FT      /note= "cytoplasmic domain"
FT      91
FT      /note= "N-glycosylated"
FT      95
FT      /note= "N-glycosylated"
FT      167
FT      /note= "N-glycosylated"
FT      258
FT      /label= Asn_Ser
FT      /note= "encoded by ART (polymorphism in nucleotide
FT      sequence"
FT      265
FT      /note= "N-glycosylated"
FT      301
FT      /label= Met_Thr
FT      /note= "encoded by AVG (polymorphism in nucleotide
FT      sequence)"

```

```

FT      Modified-site 306
FT      /note= "N-glycosylated"
FT      Modified-site 337
FT      /note= "N-glycosylated"
FT      Modified-site 361
FT      /note= "N-glycosylated"
FT      Misc-difference 447
FT      /note= "encoded by CAR"
FT      Modified-site 457
FT      /note= "N-glycosylated"
FT      Modified-site 480
FT      /note= "N-glycosylated"
FT      Modified-site 499
FT      /note= "N-glycosylated"
XX      W09803551-A1.
XX      29-JAN-1998.
XX      17-JUL-1997; 97WO-US12511.
XX      18-JUL-1996; 96US-0684594.
XX      (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX      (UTDU-) UNIV DUKE.
XX      Aruffo A, Bowen MA, Haynes BF, Margardt H, Patel D;
XX      Sladak AW.
XX      WPI: 1998-120703/11.
XX      N-PSDB: AAV13954.
XX      Claim 3; Fig 29; 130pp; English.
XX      This polypeptide is a human CD6 ligand (see AAW47088), designated
XX      activated leukocyte-cell adhesion molecule (ALCAM), that is present
XX      on the surface of chymic epithelial cells, monocytes, activated
XX      T cells and a variety of other cells. Its amino acid sequence was
XX      deduced from cDNA clones (see AAV13954) isolated from human T cell
XX      and HL60 cDNA libraries. It is the human homologue of chicken
XX      BEN, and shows homology to neuroilin, RAGE and MUC18. Anti-CD6
XX      ligand antibodies, or their binding fragments, are useful for
XX      inhibiting binding of CD6 present on the surface of a first cell
XX      to that present on a second cell. CD6 ligands and anti-CD6 ligand
XX      antibodies are also useful in screening test compounds for the
XX      ability to inhibit binding of CD6 ligand to an anti-CD6 ligand
XX      antibody (all claimed).
XX      Sequence 583 AA:
SQ
Query Match 14.6%; Score 492.5; DB 19; Length 583;
Best Local Similarity 24.8%; Pred. No. 2,7e-26;
Matches 145; Conservative 121; Mismatches 256; Indels 63; Gaps 23;
Oy      41 GSTALLKCGLSQSGNLSHVDVFSVHKEKRTILRRVROGQGS-----EPGEYERLSLOD 96
Db      36 gdtliipcrldwpg-nlmfgkwyekpgdgsyfaifstskksyqyddvpeykdrlnlse 94
Oy      97 KCATLALQVTPQDERITLCO-GKRPRSGEYRIQLRYVKAPEEPNIGNVPLGIVNSKEP 155
Db      95 -nytlisnarisdskrtvcmalvtednvfeapcltvktykpskpeivskalfi--eteq1 151
Oy      156 EVATVCGRMGYPPIQVITWYNGRPLKEEKNRVHIQSQTVES-SGLYTLQSTILKAOLVK 214
Db      152 kllgdcisedspdgnticwyrngkvlhpllegavvliifkempdvrcqlytmstleykctk 211
Oy      215 EKDQOFCELNRLPSGNHMKESREVTVPVFPPEKWLVEY-EPVGNLKGDRVEIRCL 273

```

Seq	Sequence	570 AA:
Query Match	13.9%: Score 466; DB 17; Length 570;	
Best Local Similarity	24.4%: Pred. No. 1.9e-24;	
Matches 143; Conservative 109; Mismatches 257; Indels 76; Gaps 22;		
OY	41 GSTALLKCGLSOSQNLSHVDMFVSHKEKRKLIFPVROGQGS-----EPGEYGRRLSID 96	
DB	36 gdtllpcriidvpq-nlmegkwkyekpdsprflaifirstkksayqddvpeykorlnlse 94	
OY	97 RGATLALQVTPDDEKFLPQ-GKPRROSEYRIQRYKAEDEPNICVNPGLIPVNSKE 155	
DB	95 -nyllstsnarisdetrifcmlytvednvtleaprlkvtkqspkpeylskafll--eteql 151	
OY	156 EEVATVCGRNGRPPIQVIVYKNGRPKEEKKNVHIQSSQYVES-SGLYTLQSLIKADLV 214	
DB	152 kllgdcsiedspogntwlyngltvhlplegavvllfkkmprvqlylmsclleytk 211	
OY	215 EDKRAQFYCELVNRLPSGNMKESRERVVPVPTPEKVLVEY-EPVGMLEKGDVEYICRL 273	
DB	212 adimptlcsvtylpgpsgqkllhseqavddiyyptleqvrlqlppknavkegdmltkcl 271	
OY	274 ADGNPPRH--FSISKONPSTREAEBEITNDNCVLYLEPARKHSGRYECQGLDIDPMIS 330	
DB	272 ggnupppfeelflylpgpgeqrlrsntylimd-----vrtnaqdkcslidkksmla 323	
OY	331 LLSPEQQLLVYVSDVAVSPAAP-EROEGSLTLTCAESSODLEFQWLREBTQVLERG 389	
DB	324 sta-----ltvnyl-dlslnpsgevtqigdalpvscctiasrnaivwmkdnllr--lrs 376	
OY	390 PVLQHLHDKREAGGGRVCASVSPISPOLANTQLVNVAIRGPPMMAFKERKVVWKENMVLN 449	
DB	377 p--stsslhygdaagnycetalqeveglkreslctlliveqk-p-qikmtkkcdpsgljkt 432	
OY	450 LSCGASGHPRPPTISMVNWGTAS--EEDQDDPQ--RYLSTLVNLYVTPLELTGVECTASN 503	
DB	433 lchvegipkpraigqtltsigsgvsnqnceespylmgtryss--kllispeenvt-ltctaen 489	
OY	504 DLGKNTSLILELVNLTLPDSTNTTGLSTIRASPTIRANSSTERKALPEPESGCVIV 563	
DB	490 ql-----ertvsnlvsanesrekvn-----dqakllv 517	
OY	564 AVIVICILVLAVALVAVFLY-KKGLPCRRSGKQKITLPPSRKSE 607	
DB	518 glavglllaalvagvyvwlymkkektsakhvknkdlgmteentkle 562	
RESULT 14		
ID	AABA3943	
AC	AABA3943 standard; Protein: 394 AA.	
XX		
DT	08-FEB-2001 (first entry)	
XX		
DE	Human cancer associated protein sequence SFO ID NO.1388.	
XX		
KW	Human; cancer associated gene; cancer antigen; detection; cancer;	
KW	diagnosis; cytostatic; proliferative; vulnerability; immunomodulator;	
KW	antidiabetic; antihistaminic; antirheumatic; antiarthritic; antiviral;	
KW	antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac;	
KW	dermatological; neuroprotective; thrombolytic; coagulant; neotropic;	
KW	vasotropic; antiproliferative; angiogenic; gene therapy; inflammation;	
KW	immune disorder; haematopoietic cell disorder; autoimmune disorder;	
KW	allergic reaction; graft versus host disease; organ rejection;	
KW	haemostatic; thrombolytic; cardiovascular disorder; infection;	
KW	neurological disease; drug screening.	
XX		
OS	Homo sapiens.	
XX		
PN	WO20005350-A1.	
XX		

	Query Match	12.3%	Score 414.5	DB 21	Length 394
	Best Local Similarity	29.1%	Pred. No. 4.9e-21		
	Matches 130	Conservative 67	Mismatches 170	Indels 79	Gaps 18
QY	221 FYCELNRLPLSGNHMK--ESREVTVPPEVPTDEV--WL---EVEPOMLKEGDRIETRLAD	275			
Db	1 fhxahyslpdghgldetfnllchlyptelnvqfwagspsptsgwrsrdgtvqlordg 60				
QY	276 GNPPPHSISKONPSGREAEEETTNDN--GVLYLEPAKREGSRGEQGLDLMISLT-L 332				
Db	61 gspspelylfr-----lqdegeevlnvnlegnltlegvrqgsygyrvedyaadvaql 116				
QY	333 SEPOELVNVSVDVRSPAAPEROG-----SLTITCEASSODLEFQWLRREETQ 384				
Db	117 skletelrvaylpleps-----egkalslplsasavncsvngjrlptalrytkast-- 167				
QY	385 VLERGFVLQHLDKREAGGYRCVAVSYSIPLGNRTOLVNNALFGPPMAFKRK----- 439				
Db	168 plgdgmmlaslsttidsngtycceslprtvylsrtnflvlvgsspelktaelpkadg 227				
QY	440 VWKKEMVLNLSCAASHPRPTISMNVNGTASEODODPOR---VLSTLVNLYVELLENG 496				
Db	228 sw-regdevtlcasarjnpdpklswsqslg-spspeplpgpqgwssslclktvtalsatrdg 285				
QY	497 VECTASNDLGKMTSLFLFELVNLTLTPDSVTTTGLSTASPHTEANSTGERKLPEPE 556				
Db	286 lscceasn-----phnqnkhvfifgfvspqt-----s 311				
QY	557 SRGVIVAVICILAVIGLVAFELYKKKGKLCPRSGRKEITLPSPRSSELVVEYKSOK 616				

Db	312	gagvaamavavsgllllvvavfycvirkgg-pccrgrtrekgapp---gepglshgse	367
Oy	617	LPEENGSLDQ--SSGDKRRAGDDGGEK 640 : : :	
Db	368	gpegqglimgagsgargsggfgdye 393	
RESULT 15			
ABBI0297	ABBI0297		
XX	ID	ABBI0297 standard; Protein: 749 AA.	
XX	AC	ABBI0297;	
XX	DE	10-JAN-2002 (first entry)	
XX	XX	Human cDNA SEQ ID NO: 605.	
XX	OS	Homo sapiens.	
XX	PN	WO200154474-A2.	
XX	PD	02-AUG-2001.	
PF	17-JAN-2001;	2001WO-US01349.	
XX	XX		
PR	31-JAN-2000;	2000US-179065P.	
PR	04-FEB-2000;	2000US-180628P.	
PR	24-FEB-2000;	2000US-184664P.	
PR	02-MAR-2000;	2000US-186350P.	
PR	16-MAR-2000;	2000US-189874P.	
PR	17-MAR-2000;	2000US-190076P.	
PR	18-APR-2000;	2000US-198123P.	
PR	19-MAY-2000;	2000US-20515P.	
PR	07-JUN-2000;	2000US-209467P.	
PR	28-JUN-2000;	2000US-214868P.	
PR	30-JUN-2000;	2000US-215135P.	
PR	07-JUL-2000;	2000US-216647P.	
PR	07-JUL-2000;	2000US-216880P.	
PR	11-JUL-2000;	2000US-217487P.	
PR	11-JUL-2000;	2000US-217496P.	
PR	14-JUL-2000;	2000US-218290P.	
PR	26-JUL-2000;	2000US-220963P.	
PR	26-JUL-2000;	2000US-220964P.	
PR	14-AUG-2000;	2000US-224518P.	
PR	14-AUG-2000;	2000US-224519P.	
PR	14-AUG-2000;	2000US-225213P.	
PR	14-AUG-2000;	2000US-225214P.	
PR	14-AUG-2000;	2000US-225267P.	
PR	14-AUG-2000;	2000US-225268P.	
PR	14-AUG-2000;	2000US-225268P.	
PR	14-AUG-2000;	2000US-225270P.	
PR	14-AUG-2000;	2000US-225447P.	
PR	14-AUG-2000;	2000US-225757P.	
PR	14-AUG-2000;	2000US-225758P.	
PR	14-AUG-2000;	2000US-225759P.	
PR	18-AUG-2000;	2000US-226279P.	
PR	22-AUG-2000;	2000US-226611P.	
PR	22-AUG-2000;	2000US-226688P.	
PR	22-AUG-2000;	2000US-227182P.	
PR	23-AUG-2000;	2000US-227009P.	
PR	30-AUG-2000;	2000US-228924P.	
PR	01-SEP-2000;	2000US-229287P.	
PR	01-SEP-2000;	2000US-229343P.	
PR	01-SEP-2000;	2000US-229344P.	
PR	05-SEP-2000;	2000US-229345P.	
PR	05-SEP-2000;	2000US-229513P.	
PR	06-SEP-2000;	2000US-230437P.	


```

Db 297 igsdvfltcwvgnpplltlwtwkkdsnmgprppgsppaalasqvlsnsnqllkvtga 356
QY 401 AGGGRVAVSVPSIFGLNRTQLVNVAIIFGPPMAFKERKWKYKENVYLNLSCEASGHPRP 460
Db 357 dagtlycraivprlgyaere--vplyvngpplisseavqyavrgdgg-kvecfisgtpdp 413
QY 461 ---TISMNVN---GTAS---EQDQDPQVULSTLNVLPET-LETGVECTASNDLGKN 508
Db 414 dilaawkenflevgfleytvertnsgsvstltinymeadfqlhynctawnsfgpg 473
QY 509 TSILFELVNLTTLPDSTNTTGLSTSTASPHTRANSTSTERK-LPEPESRGVIVAVI- 566
Db 474 tailgle-----erevlpv---gliagatiq 496
QY 567 VCLIVLAVLCAVLYFYKKGKLPCCRSGKOEITLPPSRKSELVE 611
Db 497 asilllffilaivfflyir-----rkgsarkdvcl---rkldikve 533

```

Search completed: June 28, 2002, 10:45:49
 Job time: 384 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2002, 10:43:45 ; Search time 18.72 Seconds

(without alignments)
842.892 Million cell updates/sec

Title: US-09-653-961-2

Sequence: 1 MGLPRIVNCAFLIAACCCPR.....SSGDKRAPDGGEXYIDLRH 646

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Match	length	ID	Description
No.	Score	%			
1	2302.5	68.5	486	2	US-08-432-016-6
2	2302.5	68.5	486	2	US-08-684-594-6
3	492.5	14.6	583	2	US-08-432-016-2
4	492.5	14.6	583	2	US-08-684-594-2
5	444.5	13.2	477	2	US-08-432-016-3
6	444.5	13.2	477	2	US-08-684-594-3
7	414.5	12.3	466	2	US-08-432-016-4
8	414.5	12.3	466	2	US-08-684-594-4
9	252	7.5	1381	4	US-09-540-245A-16
10	244.5	7.3	647	5	PCT-US93-00031-23
11	237.5	7.1	736	5	PCT-US93-00031-15
12	233	6.9	647	4	US-09-009-490A-91
13	233	6.9	647	4	US-08-482-073-5
14	233	6.9	647	4	PCT-US93-00031-11
15	232.5	6.9	1260	4	US-08-506-296B-21
16	231.5	6.9	828	1	US-08-261-304-2
17	230	6.8	735	5	PCT-US93-00031-13
18	230	6.8	739	6	US-08-482-073-6
19	230	6.8	739	5	PCT-US93-00031-9
20	226.5	6.7	662	1	US-08-261-304-7
21	226	6.7	738	3	US-08-478-208-32
22	225.5	6.7	1253	4	US-08-506-296B-14
23	225	6.7	698	2	US-08-602-725-36
24	225	6.7	734	2	US-08-389-459A-17
25	225	6.7	734	4	US-08-987-867A-17
26	224.5	6.6	1651	4	US-09-540-245A-18
27	223	6.6	738	6	Patent No. 5264554-2

28	222.5	6.6	644	5	PCT-US93-00031-21	Sequence 21, Appl
29	220.5	6.6	740	5	PCT-US93-00031-17	Sequence 17, Appl
30	220.5	6.6	1241	4	US-09-040-774-2	Sequence 2, Appl
31	219	6.5	642	1	US-08-217-299-1	Sequence 1, Appl
32	218.5	6.5	318	5	US-08-633-148-4	Sequence 4, Appl
33	218.5	6.5	340	2	US-08-633-148-2	Sequence 2, Appl
34	216	6.4	630	2	US-08-752-307B-14	Sequence 14, Appl
35	215	6.4	643	5	PCT-US93-00031-19	Sequence 19, Appl
36	215	6.4	1268	4	US-08-506-296B-28	Sequence 28, Appl
37	213.5	6.3	1356	1	US-08-810-116-8	Sequence 8, Appl
38	213.5	6.3	1356	2	US-07-930-548A-8	Sequence 8, Appl
39	212	6.3	1395	4	US-09-540-245A-15	Sequence 15, Appl
40	207	6.2	788	1	US-08-232-538-15	Sequence 15, Appl
41	207	6.2	788	2	US-08-786-164-15	Sequence 15, Appl
42	205.5	6.1	1356	4	US-09-098-707A-2	Sequence 2, Appl
43	205.5	6.1	1356	4	US-09-483-539-2	Sequence 2, Appl
44	204	6.1	612	2	US-08-752-307B-11	Sequence 11, Appl
45	203.5	6.1	1266	4	US-08-506-296B-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-432-016-6
Sequence 6, Application US/08432016
Patent No. 5968768
GENERAL INFORMATION:
APPLICANT: HAYNES, BARTON F.
APPLICANT: ARUFFO, ALEANDRO
APPLICANT: PATEL, DHAVALKUMAR
APPLICANT: BOWEN, MICHAEL A.
APPLICANT: MARQUARDT, HANS
TITLE OF INVENTION: CD6 LIGAND
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHIVE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/432,016
FILING DATE: 01-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/333,350
FILING DATE: 02-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/143,903
FILING DATE: 02-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: WILSON, MARY J.
REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 1579-95
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-432-016-6

Query Match 68.5%; Score 2302.5; DB 2; Length 486;
 Best Local Similarity 91.5%; Pred. No. 2,3e-178;
 Matches 450; Conservative 9; Mismatches 22; Indels 11; Gaps 3;

QY 40 VGSTALLKCGLSQSGCNLSHVDWFSVHKERKRLIFRVNQGQSGSEGEYEQRLSLQDRGA 99
 ||||||| : : : : :
 DB 1 VGSTALLKCGLSQSGCNLSHVDWFSVHKERKRLIFRVNQGQSGSEGEYEQRLSLQDRGA 60

QY 100 TLALQVTPDDEIFLFCQGR--PR---SQEYRIQLRYKAPDEEPIQVNPGLGIPVNSKE 154
 ||||||| : : : : :
 DB 61 TLALQVTPDDEIFLFCQGRKRGPRSTASSASTKLRL-----PNIQVNPGLGIPVNSKE 114

QY 155 PEEVATCVGRNGYPIPOVYIWKNGRPLKEEKNRVHIQSSQIVESSGLTYLQSLKAQLVK 214
 ||||||| : : : : :
 DB 115 PEEVATCVGRNGYPIPOVYIWKNGRPLKEEKNRVHIQSSQIVESSGLTYLQSLKAQLVK 174

QY 215 EDKDAQFYCELNRYLPSGNHMKESREVTVPVFPYEKWLVEPEVGMLEKEDRVEIRCLA 274
 ||||||| : : : : :
 DB 175 EDKDAQFYCELNRYLPSGNHMKESREVTVPVFPYEKWLVEPEVGMLEKEDRVEIRCLA 234

QY 275 DGNPPPHFSISKONPSTREAEETTNDNGVLYLEPARKHSGRYECQGLDIDTMTISLSE 334
 ||||||| : : : : :
 DB 235 DGNPPPHFSISKONPSTREAEETTNDNGVLYLEPARKHSGRYECQGLDIDTMTISLSE 294

QY 335 POELLVNVSDVRSVPAAPERQEGSSLTLTCEASSODLEFQWLREETGVLERGPVQL 394
 ||||||| : : : : :
 DB 295 POELLVNVSDVRSVPAAPERQEGSSLTLTCEASSODLEFQWLREETGVLERGPVQL 354

QY 395 HDLKRAGGGRYCVASVSPISGLNRTQLVNVAIFGPPMMAFERKRYWKENVNLSCGA 454
 ||||||| : : : : :
 DB 335 HDLKRAGGGRYCVASVSPISGLNRTQLVNVAIFGPPMMAFERKRYWKENVNLSCGA 414

QY 455 SGHPRPTISMVNGTASEDQDOPQRYLSTLVNLYTPPELLETGVECTASNDLGKNTSILFL 514
 ||||||| : : : : :
 DB 415 SGHPRPTISMVNGTASEDQDOPQRYLSTLVNLYTPPELLETGVECTASNDLGKNTSILFL 474

QY 515 ELVNLTTLPDS 526
 ||||||| : : : : :
 DB 475 ELVNLTTLPDS 486

RESULT 2
 US-08-684-594-6
 ; Sequence 6, Application US/08684594
 ; Patent No. 5998172
 ; GENERAL INFORMATION:
 ; APPLICANT: HAYNES, BARTON F.
 ; APPLICANT: ARUFFO, ALEJANDRO
 ; APPLICANT: PATEL, DHAVALKUMAR
 ; APPLICANT: BOWEN, MICHAEL A.
 ; TITLE OF INVENTION: CD6 LIGAND
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: NIXON & VANDERHAYE P. C.
 ; STREET: 1100 NORTH GLEBE ROAD
 ; CITY: ARLINGTON
 ; STATE: VIRGINIA
 ; COUNTRY: U.S.A.
 ; ZIP: 22201-4714
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/684,594
 ; FILING DATE: 18-JUL-1996
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/432,016
 ; FILING DATE: 01-MAY-1995

PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/333,350
 ; FILING DATE: 02-NOV-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/143,903
 ; FILING DATE: 02-NOV-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: WILSON, MARY J.
 ; REGISTRATION NUMBER: 32,955
 ; REFERENCE/DOCKET NUMBER: 1579-112
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 816-4000
 ; TELEFAX: (703) 816-4100
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 486 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-684-594-6

Query Match 68.5%; Score 2302.5; DB 2; Length 486;
 Best Local Similarity 91.5%; Pred. No. 2,3e-178;
 Matches 450; Conservative 9; Mismatches 22; Indels 11; Gaps 3;

QY 40 VGSTALLKCGLSQSGCNLSHVDWFSVHKERKRLIFRVNQGQSGSEGEYEQRLSLQDRGA 99
 ||||||| : : : : :
 DB 1 VGSTALLKCGLSQSGCNLSHVDWFSVHKERKRLIFRVNQGQSGSEGEYEQRLSLQDRGA 60

QY 100 TLALQVTPDDEIFLFCQGR--PR---SQEYRIQLRYKAPDEEPIQVNPGLGIPVNSKE 154
 ||||||| : : : : :
 DB 61 TLALQVTPDDEIFLFCQGRKRGPRSTASSASTKLRL-----PNIQVNPGLGIPVNSKE 114

QY 155 PEEVATCVGRNGYPIPOVYIWKNGRPLKEEKNRVHIQSSQIVESSGLTYLQSLKAQLVK 214
 ||||||| : : : : :
 DB 115 PEEVATCVGRNGYPIPOVYIWKNGRPLKEEKNRVHIQSSQIVESSGLTYLQSLKAQLVK 174

QY 215 EDKDAQFYCELNRYLPSGNHMKESREVTVPVFPYEKWLVEPEVGMLEKEDRVEIRCLA 274
 ||||||| : : : : :
 DB 175 EDKDAQFYCELNRYLPSGNHMKESREVTVPVFPYEKWLVEPEVGMLEKEDRVEIRCLA 234

QY 275 DGNPPPHFSISKONPSTREAEETTNDNGVLYLEPARKHSGRYECQGLDIDTMTISLSE 334
 ||||||| : : : : :
 DB 235 DGNPPPHFSISKONPSTREAEETTNDNGVLYLEPARKHSGRYECQGLDIDTMTISLSE 294

QY 335 POELLVNVSDVRSVPAAPERQEGSSLTLTCEASSODLEFQWLREETGVLERGPVQL 394
 ||||||| : : : : :
 DB 295 POELLVNVSDVRSVPAAPERQEGSSLTLTCEASSODLEFQWLREETGVLERGPVQL 354

QY 395 HDLKRAGGGRYCVASVSPISGLNRTQLVNVAIFGPPMMAFERKRYWKENVNLSCGA 454
 ||||||| : : : : :
 DB 335 HDLKRAGGGRYCVASVSPISGLNRTQLVNVAIFGPPMMAFERKRYWKENVNLSCGA 414

QY 455 SGHPRPTISMVNGTASEDQDOPQRYLSTLVNLYTPPELLETGVECTASNDLGKNTSILFL 514
 ||||||| : : : : :
 DB 415 SGHPRPTISMVNGTASEDQDOPQRYLSTLVNLYTPPELLETGVECTASNDLGKNTSILFL 474

QY 515 ELVNLTTLPDS 526
 ||||||| : : : : :
 DB 475 ELVNLTTLPDS 486

RESULT 3
 US-08-432-016-2
 ; Sequence 2, Application US/08432016
 ; Patent No. 5968768
 ; GENERAL INFORMATION:
 ; APPLICANT: HAYNES, BARTON F.
 ; APPLICANT: ARUFFO, ALEJANDRO
 ; APPLICANT: PATEL, DHAVALKUMAR
 ; APPLICANT: BOWEN, MICHAEL A.

Db	173	LQYARKEDEVSOQTC-----TAKHYMBGDQVSEPESEPIHHPFEKVALQVWOSOPIR	225
Qy	264	EGDRVEIRCLADGNPP--HEFISKONPSTREAEETNDNGVLYLEPAKHEHSRYEC	320
Db	226	EGEDVTLTCQADGNPPPTSEFNFI-----KKKRYTVDKDYTLTGVTGRADSGIYKC	277
Qy	321	QGLDLPFISLSEPOELLVYVDVWVSPAPE--RQEGSLITVTCBAESSDLEFOWLR	379
Db	278	SLLDNDVWES-----TQFYWSFL-DVSLTPGKLYKLVNGEHLVSLDKNASSSEKRYWTK	332
Qy	380	EETGOVLERGEFVLQJLHDIKREAGSGGYRCVAVSPSIPGLNFTQLVNVAIFGPPWMA--EKE	437
Db	333	D--NNKIDKLP--DESKLTYS DAGLYCDV---SIEGIKRSLSPELIYEGIPKTTSLTKH	385
Qy	438	RKVVWKKEMVVLNLSCEASGHRPPIITSMWVKNSTABEQDDQPRVSLTINLVATPELLETGV	497
Db	386	RSSDQKHRY--LTCCEAGSPKPDVOWSVNCTNDEVSYNNKATYKLTLYVPSKNLT--V	439
Qy	498	ECTASNDGKNT	509
Db	440	SCLYTNKIGEDT	451

```

;      TOPOLOGY: linear
;      MOLECULE TYPE: peptide
US-08-684-594-4

```

4.5; DB 5; Length 647;
0.1.8e-11;

atches 255; Indels 181; Gaps 32;

FSVHKEKRTLIFRVRGQGQSEPGYEQ 90

```

|| | : : ||
-----RTQIDSP LSGKVRSE----- 74

```

RPRSOEYRIOLBVKAPPEPNIQVN-PT. 146

```

: | | : | | : | | : |
- - K T E K C I O V E L Y C E B K D P E T I U S C O N T 1 2 3

```

ANLENGIÖVELISTFUNDPEIHLISGPL 12/

KNGRPLKEKNRVHIQSSQTVESSGL-Y 202
| : : : :: | : |

LLKGDHLMKSQEFLEADADRKSLKTSLEV 180

MMKESREVT--VPVFYPTKEVWLEVEPVG 260

VPTVRQAVKELQVYISPKNVTVISVNPST 234

EAEEETJNDNGVLLEPARKEHSGRYEC 320

```

      : | | | | | |
NGNLQHLSGNATLTLIAMRMEDSGIYVC 291

```

PERO-----EGSSTJTCFAES---SQ 371

TIME	MODE	Q	Q/I
:	: :: :		
DEFECC	INCC	CC	CC

FELEMSGGVNGSSVIVSCNVPVIFLD 342

```

-----REAGGRCVASY----- 411
      : | | :

```

SLEMTFIPTIEDTGKALVCQAKLHIDM 400

WVKENMVL-----NLSCEASGHPRTI 462

LVSPSSILEGSSVNMTCLSQGFAPKI 452

TCVE-----CTASNDLGKNTSILEL 516

TMEDSGVYLCEGINOAGRSRKEVELLI 504

-----PANCSTFBR----- E51
DHT-----

YCC - - - - - UNCTICNUN
| | : |
| | : |

PEI:W11LKKA:EI:GD1VLKSIDGAYTIR 564

-----IVAVICILVLAVLGAV 578
:| :| :| :

ENNKDYFPELLVYFASSLIIPAIGMI 624

NIZING THE FOURTH

LIKE DOMAIN OF VCAM1

Query Match 6.9%; Score 232.5; DB 4; Length 1260;
Best Local Similarity 22.4%; Pred. No. 4.8e-10;
Matches 143; Conservative 68; Mismatches 257; Indels 171; Gaps 28;

```
OY 25 PGEAOPAPBELVEVEVGSTALLKGLSOSQGNLSHVDFSVHKKEKRTLIIFVRQGGQSE 84
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 138 PREYVKP-----VEVEGESVLPFCNPPPSAAP-PRITYMN-----SKIFDIKO----- 180
OY 85 PGEYBQRLSLQDRGATLALTOVTPODERIFLQ-----GKRPSOEYRIQLRVYKAPEEP 140
    : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 181 ---DERVSMGNGDLYFANVLTSNDHSDYICNAHPGTRIIQKEPIDLAV--KPTNSM 234
OY 141 IOVNP-LGIPIVNSKE-----PEEVATCVGRNGYPIPIQVIWYKNGRPLKEEKNRHIO 192
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 235 IDRKPRLLFPTNSSSRLLVALOGOSLILECIA-EGFPPTIKWLHPSDMPD--RVIYON 291
OY 193 SQTVESSGLYTIOSILKAQLKEDKDAQFCELVNRIPLSGNHMKESREVTVPVFPTEKY 252
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 292 HNK-----TIQQL-----NWGEEDGETTCLAENSLGSAKH-----AYTVVEA 330
OY 253 ---MLEVEPVGMLEKGDREVEIRCLADGNPPHFSISKQNPSTREAEEETINDGVLVLEP 309
    | | : | : | : | : | | | | | | | | | | | | | | | | | | | | | | | |
Db 331 APYMLQKPGSHLYGPGETARLDQVQGRQP-----ETWIRING-MSMET 374
OY 310 ARKEHSGRYECOGDL-----DTWISLSEPOE---LVN---YVSDRVSPAPERO- 356
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 375 VNKDQRYRIEOGSLILSNVQPTDMVTCCEARNOHGILLANAYIYVQLPARIITKDNQT 434
OY 357 ---EGSSLITLCEAESQDLEFQMLRETEGOVLE-----RGPVLQIHDKREAGGY 405
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 435 YMAVPGSTAYILLCKAFGAPVPSVQMLDEGTTVLQDEREFPPYANGTLLSRDLQANDTGRY 494
OY 406 RCVASVPSPISGLNRTQLVNVAIFGPPWMAFEKRWVWKENNVNLSCASGHP--RPTIS 463
    | | : | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 495 FCOAANDQ---NNVTILANLQVKEATOITQGPSAIEKKARVTFTCQASFPDLSQASIT 551
OY 464 WNVNCTASEQODOPORVLTANLVTPEL---LETGVECTASNDL-----GKN 508
    | : | : | : | : | : | | | | | | | | | | | | | | | | | | | | | |
Db 552 WRGDRDLQERGDSDKYPFIEDGKLVLSQSDQGNYSQVASTELDEYESRAQLLVGSP 611
OY 509 TSLFLFLVNTLTLLPDSNTTGTGSTASPHTR----- 542
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 612 GPVPHLELSDRHL---KQSQVHLMSWSPAEDHNSPIEKYDIEFEDKEAPKWFSLGKVP 668
OY 543 ANSTISTERKL-----PEPESRGVY 561
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 669 GNQSTSTLKLSPYHYTFRTAINKYGGPESPYSESIV 707
```

Search completed: June 28, 2002, 10:46:59
Job time: 194 sec

QY 241 VTVPVFPTEKVMLEVPVGMKKGDRVEIRCLADGNPPPHFSTSKONPSTREAEETTN 300
Db 241 VTVPVFPTEKVMLEVPVGMKKGDRVEIRCLADGNPPPHFSTSKONPSTREAEETTN 300
QY 301 DNGVLVEPARKESGREGCGDLDTMISLSEPEQLLVYVSDVAVSPAAPEROGSS 360
Db 301 DNGVLVEPARKESGREGCGDLDTMISLSEPEQLLVYVSDVAVSPAAPEROGSS 360
QY 361 LITLCEASSODLEFQWMLREETGOVLERGVLQHLHDKREAGGRCVAVSPSTPGINRT 420
Db 361 LITLCEASSODLEFQWMLREETGOVLERGVLQHLHDKREAGGRCVAVSPSTPGINRT 420
QY 421 QLVNVAIFGPPWMAFKERKVMKENVMLNLSCEASGHPRTISNNVNGTASSEODODORV 480
Db 421 QLVNVAIFGPPWMAFKERKVMKENVMLNLSCEASGHPRTISNNVNGTASSEODODORV 480
QY 481 LSTLVNLTVPPELTGVECTASNDLGKNTSILFELVNLTTLPDSTNTTGLSTSTASPH 540
Db 481 LSTLVNLTVPPELTGVECTASNDLGKNTSILFELVNLTTLPDSTNTTGLSTSTASPH 540
QY 541 TRANSTSTERKLPESRGVIVAVIYCIYLVAVLAVLYLKKKLPCKRSGKOEITL 600
Db 541 TRANSTSTERKLPESRGVIVAVIYCIYLVAVLAVLYLKKKLPCKRSGKOEITL 600
QY 601 PPSRSELVVEVSKDLPPEEMGLQSSGDKRAPDGEKTYIDLRH 646
Db 601 PPSRSELVVEVSKDLPPEEMGLQSSGDKRAPDGEKTYIDLRH 646

RESULT 2

150419
s-glycerin precursor - chicken
C:Species: Gallus gallus (chicken)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: 150419
R:Taïra, E.; Takaha, N.; Taniura, H.; Klm, C.H.; Miki, N.
Neuron 12, 861-872, 1994
A:Title: Molecular cloning and functional expression of glicerin, a novel cell adhesion p
A:Reference number: 150419; MUID:94213753
A:Accession: 150419
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-584 <TAI>
A:Cross-references: GB:D38559; NID:g1009246; PIDN:BAA07563.1; PID:g559701

Query Match 25.7%, Score 863; DB 2; Length 584;
Best Local Similarity 35.1%, Pred. No. 2e-47;
Matches 216; Conservative 107; Mismatches 210; Indels 82; Gaps 19;
QY 8 CAFLAACCCCPVAGVGEAOPAPDELVEVSGTALLKGL-----SOSQGNLSHVDW 62
Db 22 CPEFLLLCC-----GAAGR-----EYMPSSA-----GSGDRHSQAGVOLLH-PW 63
QY 63 FSVHKEKRTLLFR---VROGOG-----OSEPEYEOURLSDRGATLALTVQTPDE 111
Db 64 ECLLLRLRVLCQPRPLASGEAVRNHAGVRIDETSEYERLSVGRDKA-LISKVTQDN 122
QY 112 -RIPLCQ--GKRPRSQERYIQLRYTKARBEPIQVNPJGIPYNSKEPEVATCGRNGCP 168
Db 123 ARTFCQVADSGVSGESTELTYTKIPAPPEITPNSAGIPQOSNDMKIAQCTSENSEFP 182
QY 169 IPQVITWYKNGRPLKEKNVHIQSSQOTVSSGLYTLQSLTAQVLEKEDKQAFCELYNR 228
Db 183 SNITWYKNGRPLKEEDTKLITTLVRESNGLYTVSTLSKVIRERNSLFRCTVHW 242
QY 229 LPSGNHMKSRREVTVPVFPTEKVMLEVP--VGMKKGDRVEIRCLADGNPPPHFSTSKO 287
Db 243 LOGQRTKDSRPVNTVFPTEKVMLEVP--VGMKKGDRVEIRCLADGNPPPHFSTSKO 302
QY 288 --NPSTREAEETTNNGVLVEPARKESGREGCGDLDTMISLSEPEQLLVYVSD 345
Db 288 --NPSTREAEETTNNGVLVEPARKESGREGCGDLDTMISLSEPEQLLVYVSD 345

Db 303 ELGDSWQDMTSLADTNDGVLMHNVSKESSGLYRCQITLDDMTQHEGD-VELVYNTIEG 361
QY 346 VRY--SPAAPERQESSLITLCEASSODLEFQWMLREETGOVLERGVLQHLHDKREAGG 403
Db 362 VQVKMPPSSP-LHEGDSVLSCTASHSPVKLDYQ-RDARGKRVABGNOLLTNLTFEISS 419
QY 404 GYRCVAVSPISPIQNTQLVNVAIFGPPWMAFKERKVMKENVMLNLSCEASGHPRTIS 463
Db 420 NFSCRVKARSVPELEQSKQVAVAKKPRIVALSAPLYVRQDEVLNLCOKAIAFPQSPH 479
QY 464 WNVNGTASSEODDOPQVYVSTLVNLTVPPELTGVECTASNDLGKNTSILFELVNLTT 523
Db 480 WSYNGTTHY--MENQHMASNLTVRVSHDLIRAGACORVSNALGVSE----KHQLDQK 533
QY 524 PDSNTTGLSTSTASPHSTANSTSTERKLPESRGVIVAVIYCIYLVAVLAVLYLKKK 583
Db 534 PS-----ESKQIITVAVIYCIYLVAVLAVLYLKKK 563
QY 584 KKGKLPCKRSGKOEI 598
Db 564 KKGKISGSGRGKQDI 578

RESULT 3

138000
Lutheran blood group glycoprotein precursor - human
C:Species: Homo sapiens (man)
C:Date: 09-Mar-1996 #sequence_revision 09-Mar-1996 #text_change 05-Nov-1999
C:Accession: I38000; S11663
R:Parsons, S.F.; Mallinson, G.; Holmes, C.H.; Houlihan, J.M.; Simpson, K.L.; Mawby, W
Proc. Natl. Acad. Sci. U.S.A. 92, 5496-5500, 1995
A:Title: The Lutheran blood group glycoprotein, another member of the immunoglobulin
A:Reference number: I38000; MUID:95296337
A:Accession: I38000
A:Molecule type: mRNA
A:Residues: 1-628 <RES>
A:Cross-references: EMBL:X83425; NID:g603559; PIDN:CA56449.1; PID:g603560
A:Note: parts of this sequence, including the amino end of the mature form, were conf
A:Genetics:
A:Gene: GDB:LU
A:Cross-references: GDB:120155; OMIM:111200
A:Map position: 19q12-19q13
A:Keywords: glycoprotein
F:1-31/Domain: signal sequence #status predicted <SIG>
F:32-628/Product: Lutheran blood group glycoprotein #status experimental <MAY>

Query Match 19.3%, Score 647.5; DB 2; Length 628;
Best Local Similarity 29.5%, Pred. No. 1.2e-33;
Matches 202; Conservative 103; Mismatches 265; Indels 115; Gaps 27;
QY 2 GLPR-LVCAFLAACCCCPVAGVGEAOPAPDELVEVSGTALLKGLSOSQGNLSH- 59
Db 12 GARPLLLAVLLA-----HPDAQAEVRLSVPRIVEVMKSGKVLIDC---PPTGNDHY 62
QY 60 -VDNF-----SVHKEKRTLLFRVROGOGSSEPEYEQRLSLDRGATLALTVQ 106
Db 63 MLEWFLIDRSGARPLASAMQSELDQVTMHDHGRGSPR---YQDSQGR--LVLA 115
QY 107 TPQDERFLC--GCKRPRSQERYIQLRYTKARBEPIQVNPJGIPYNSKEPEVATCGRNGCP 164
Db 116 QVGERIYVCVVRGAAGTAEATRLNVEFAKPRATEVSPKGLTSLWEDSADQITACNSR 175
QY 165 NGYIPQVITWYKNGRPL--EKNRVHIQSSQTV-ESSGLYTLQSLTAQVLEKEDKQAFCE 221
Db 176 NGNAPRITWYRNQRLVEVENKPEGYMSKRVIRASGSLISLTVLRLRKDDQASR 235
QY 222 YCELNRYLPSGNHMK--BSREVTVPVFPTEKVMLEVP--VGMKKGDRVEIRCLADG 276
Db 236 HCAHAYSLPRRHRQRLDSPFHLLHYTRDHYQFWGSPSTPAGWVREGDPTVOLLRCGSG 295
QY 277 NPPHSTISKONPSTREAEETTNNDN--GVLEPARKESGREGCGDLDTMISLSEPEQLLVY 333
Db 277 NPPHSTISKONPSTREAEETTNNDN--GVLEPARKESGREGCGDLDTMISLSEPEQLLVY 333

Db 296 SPSEYTLFR-----LQDEQEEVLNVNLEGNLTLEGVTRGSGTGYCQVEDYDAADVDYLS 351

QY 334 EPQELLVNVSDVRSVPAAPERQEG-----SSLTITCEASSQDLFQMLREBTQV 385

Db 352 KTELRLVAVLDPLELS-----BEKVLSLPLNSAVVNCVHGILPTPALMTKOST--P 402

QY 386 LERGVQLDHLKREAGGRCVAVSPSIPGLNRTOLVNAIFGPPMAFKERK-----V 440

Db 403 LGDGMPLSLSTITFDNSNGTYCEASLPTVPVLSRTQNTFLVQSGPELKTAEIPEKADGS 462

QY 441 WKEMNVLNLSCEASGCHRPITISMNVNCTASEQDQDPOR---VLSTLNVLTPELLETGV 497

Db 463 W-REGEDEVTLICSGARGHPDKLSWSQLG-GSPAEPFGRGWSSSLTKVTSALSROGI 520

QY 498 ECTASNDLGKNTSILPLELVNLTTLTPDSTNTTGLSTASPHTRANSTERTKLPPEPS 557

Db 521 SCEASNPHGNKRHVHFGAAS-----PQTSQ 546

QY 558 RGVIVAVIYICILVAVLGAVALYFLYKKGKLPGRSGKQETLTPPSRKSSELVVEVKKDKL 617

Db 547 AGVAVMAVAVSVGLLLLVAVAFYCVRRKGG-PCCRQRRKRGAPP---GEPLSHSGSEQ 602

QY 618 PEEMGLQGG--SSGDKRPRGQGER 640

Db 603 PEQGTGLMGASGARGSGSGFGDE 627

RESULT 4

B-CAM protein - human
C:Species: Homo sapiens (man)
C>Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 21-Jul-2000
C/Accession: I37202; S47272
R:Campbell, I.G.; Foulkes, M.D.; Senger, G.; Trowsdale, J.; Garin-Chesa, P.; Rettig, W.J.
Cancer Res. 54, 5761-5765, 1994
A:Title: Molecular cloning of the B-CAM cell surface glycoprotein of epithelial cancers:
A:Reference number: I37202; MUID:95042297
A:Accession: I37202
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-588 <RES>
A:Cross-references: EMBL:X80026; NID:9535178; PIDN:CAA56327.1; PID:9535179
A:Genetics:
A:Gene: B-CAM

Query Match 18.0%; Score 606; DB 2; Length 588;

Best Local Similarity 29.2%; Pred. No. 5.1e-31; Mismatches 247; Indels 110; Gaps 25;

Matches 187; Conservative 96; Mismatches 247; Indels 110; Gaps 25;

QY 2 GLPR-LVCAFLAACCCCPRYAGVGEAEQAPPELVEVEGSTALLKCGLSQSGNLSH- 59

Db 12 GAPRLILLAVLLAA-----HPDAQAEVRLSVPLVEVMGKSVILDC--TPTGTHDH 62

QY 60 -VDMF-----SVAKKKTLLFRVROGQSGSEPEBYEORLSLDRGATLAIQV 106

Db 63 MLEWFLDRSGARPLASAEWQSELOVTMHDTRGRSP---YQDSQGR--LVLAFA 115

QY 107 TPOERIFLC--QGRKRPSEYRIQLRVYKAPPEPNIQVNPGLGIPVNSKPEEVAICVGR 164

Db 116 QVGEERIVYCVVRAAGAGTAATARKLVFAKPEATEVSPKKGKTLISWEDSAQELATCNSR 175

QY 165 NGYDIPQVIVYKNGRPK--EKKNRVHIQSSQTV-ESSGLYTLQSLKAQLVDEKDAQF 221

Db 176 NGNPAPKTYWRNGORLVEPEMNPESGYMSRTVREASGLSLSTLYLPCRRDDRRASF 235

QY 222 YCELNTYLPDSGNHMK-ESRETYVVFYPTKEV--WL--EVEPQMKKEGRLVEICLADG 276

Db 236 HCAAHYSIPBGRGRDLSPTFHLTLHYPTHEVQFWGSPSTPAGWAVEGDTVOLLRGDS 295

QY 277 NPPHFSISKQNPSTRAEETINDN--GVLVLEPARKHSGRGECGLDIDTMIISLSE 334

Db 296 SPSEYTLFR-----LQDEQEEVLNVNLEGNLTLEGVTRGSGTGYCQVEDYDA----- 344

QY 335 PQELLVNVSDVRSVPAAP-ERQEG-----SSLTITCEASSQDLFQMLREBTQV 385

Db 345 ADDVQSLKTLVDRAVAYADPLELSEGGKVLSTPLNSSAVVNCVHGILPTPALMTKOST--P 402

QY 386 LERGVQLDHLKREAGGRCVAVSPSIPGLNRTOLVNAIFGPPMAFKERK-----V 440

Db 403 LGDGMPLSLSTITFDNSNGTYCEASLPTVPVLSRTQNTFLVQSGPELKTAEIPEKADGS 462

QY 441 WKEMNVLNLSCEASGCHRPITISMNVNCTASEQDQDPOR---VLSTLNVLTPELLETGV 497

Db 463 W-REGEDEVTLICSGARGHPDKLSWSQLG-GSPAEPFGRGWSSSLTKVTSALSROGI 520

QY 498 ECTASNDLGKNTSILPLELVNLTTLTPDSTNTTGLSTASPHTRANSTERTKLPPEPS 557

Db 521 SCEASNPHGNKRHVHFGAAS-----PQTSQ 546

QY 558 RGVIVAVIYICILVAVLGAVALYFLYKKGKLPGRSGKQEG 597

Db 547 AGVAVMAVAVSVGLLLLVAVAFYCVRRKGG-PCCRQRRK 585

RESULT 5

JH0506
adhesion molecule SCL precursor - chicken
C:Species: Gallus gallus (chicken)
C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 21-Jul-2000
C/Accession: JH0506; PS0270
R:Tanaka, H.; Matsui, T.; Agata, A.; Tomura, M.; Kubota, I.; McFarland, K.C.; Kohr, B.
Neuron 7, 535-545, 1991
A:Title: Molecular cloning and expression of a novel adhesion molecule, SCL.
A:Reference number: JH0506; MUID:92030150
A:Accession: JH0506
A:Molecule type: mRNA
A:Residues: 1-588 <TRAN>
A:Cross-references: GB:S63276; NID:9238000; PIDN:AAB20170.1; PID:9238001
A:Experimental source: embryo
A:Accession: PS0270
A:Molecule type: protein
A:Residues: 34-48 <TRAN>
C:Comment: this protein is uniquely and transiently expressed on spinal cord motoneur
C:Keywords: glycoprotein; transmembrane protein
F:133/Domain: signal sequence #status predicted <SIG>
F:34-588/Product: adhesion molecule SCL #status predicted <ADH>
F:500-523/Domain: transmembrane #status predicted <TRA>
F:101,173,199,271,312,366,462,465,504/Binding site: carbohydrate (Asn) (covalent) #st

Query Match 15.2%; Score 509.5; DB 2; Length 588;

Best Local Similarity 26.4%; Pred. No. 7.3e-25; Mismatches 254; Indels 91; Gaps 30;

Matches 166; Conservative 117; Mismatches 254; Indels 91; Gaps 30;

QY 6 LVCAFLAACCCCPRYAGVGEAEQAPPELVEVEGSTALLKCGLSQSGNLSHYD-FS 64

Db 18 LILALCMPLALGL-----YTVNAVYGDITMPCRIELVDDG-LMGKMKYE 65

QY 65 VHKKKTLL-FR--VNOGQSGSEPEYEBORLSLDRGATLAIQVTPDERFLICQ--GK 119

Db 66 MPRGSEVYFAFSSSTKKNVQYDVPYKRLSLSE--NTLSTKNARISDEKRFVCMVTE 124

QY 120 RPRSEYRIQLRVYKAPPEPNI--QVNPGLGIPVNSKPEEVAICVGRNGYDIPQVITWKN 177

Db 125 DVVSEEPY-VVVFQKQSPETLHQADFL---ETEKLMKGECCVYRSYBEGNWTWKN 179

QY 178 GRPLK--EKKNRVHIQSSQTVSSGLYTLQSLKAQLVDEKDAQFYELNTRLPDSGNHM 235

Db 180 GRVLPQVEEVVIVINLRKEN-RSTGLFTMTSSIQYMPKREANAKFTCTIVYHGPSQGT 238

QY 236 KESRETYVVFYPTKEVWLEV--EPVGMKKEGRLVEICLADGNPPH--FSISKQNPST 291

Db 239 IQSEPVVFPVHYPTKEVITRVLVSQSTIKESGDNTVLKSGNGNPPQDFLFYI----- 291

QY 292 RRAEETINDNGVLVLEPARKHSGRGYECGLDIDTMIISLSEPQELLVNVSDVRSVPA 351

Db 292 -PGEFEGIRSSDYVTMTVRNRNATGEYKCSLIDKSM-----DNTTITVHL-DLQLPFS 344
Qy 352 AP-ERQEGSSLTITCEASSODLEFQWLREBTGOVLERGVLQHLDKREAGGGRVAVS 410
Db 345 GEVTKQIEALPVSCITSSSRNATVFWIKDNTNR--MKTSP--SFSSLOYODAGNYICETT 400
Qy 411 VPSITGLNRTQLVNVAIFGPPWMAFKERKVVWKENVNLNLSCEASGHPRTISMNVNGTA 470
Db 401 LQEVGGLKKRRTKLIVGKRP--QIKMTKKTNTNKNKSKTIVCHVEGFPPKPAQMTVIGSG 458
Qy 471 S--EQDODPQVRLSTLN--VLVTPELLETVGECTASNDLGKNTLSLELVNLTLPDPS 526
Db 459 SLINKTEETKYVNGKFSKKIITIAPEENVTL-LNCIAENEL-----E 497
Qy 527 NNTTGLSTSTAS-----PHTRANSTSTERKLPPEPSRGVIVAVYICVLAVLAGAVLY 580
Db 498 RVTSLNVSATSIPEYDEPEDR--NDNSEK----VNDQAKLIVIGVGLLVALVAGVY 552
Qy 581 FLY-KKGKLLPCRRSGKQETILPPSRKSE 607
Db 553 WLTVKSKTSKXKHVDKLDGNIENKKE 580

RESULT 6
A45254
surface glycoprotein BEN precursor - chicken
C:Species: Gallus gallus (chicken)
C>Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 21-Jul-2000
C:Accession: A45254; S19202
R:Pourquie, O.; Corbel, C.; Le Caer, J.P.; Rossier, J.; Le Douarin, N.M.
Proc. Natl. Acad. Sci. U.S.A. 89, 5261-5265, 1992
A:Title: BEN, a surface glycoprotein of the immunoglobulin superfamily, is expressed in
A:Reference number: A45254; MUID:92302224
A:Accession: A45254
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-588 <POD>
A:Cross-references: EMBL:X64301; NID:963087; PIDN:CAA45579.1; PID:963088
C:Keywords: glycoprotein

Query Match 15.0%; Score 505.5; DB 2; Length 588;
Best Local Similarity 26.4%; Pred. No. 1.3e-24;
Matches 166; Conservative 116; Mismatches 255; Indels 91; Gaps 30;
Qy 6 LVCAFLAACCCCPRVAGPEAEQAPAPELVEVEGSTALLKCGLSQSGNLSHVDW-FS 64
Db 18 LLC--LLALALCMPPALGL-----YTVNAVYGDITIMPCRLVDPG-LMFGKMYE 65
Qy 65 VHKEKRTLI-FR--VRQGGQSGEPGEYEOURLSLODRGATLALQVTPODERIFLCO--GK 119
Db 66 MNGSPVFIAPFSSSTKKNQYDVPDKDRLSLSE--NTTSLKNARISDEKRFVCMLYTE 124
Qy 120 RPRSGEYRIQLRVYKAAPEEPNI--QVNPGLIPVNSKEPEVATCYGRNGYPIPOVIWYKN 177
Db 125 DVVSEEPYV-KVYFKQSPQPEILHQADF-----ETEKLMGLGCVYRSDYPEGNVTWYKN 179
Qy 178 GRPLK--EKNRVHIQSSQTVESGLTYLQSLKQALVQKDAFQCELANRLPSGNHM 235
Db 180 GRVLPQVEEVVAVINLKRVEN-RSTGLFTMTSSLOYMPTEKDANAKFTCIYTVHGSGOKT 238
Qy 236 KESRETVVAVPYPEKRWLEV--EPVGMLEKGGDRVEIRCLADGNPPH--FSISKONPST 291
Db 239 IOSEPVEFVHYPTKEKTVIRVLSQSSSTIKEGNVTLKCGNGNPPPOEFLFYI----- 291
Qy 292 REAEETITNDNGVILVLPARKHSGRYECQGLDITMISLSEPOELLVNYSDVRVSPA 351
Db 292 -PGETEGIRSSDYVTMTVRNRNATGEYKCSLIDKSM-----DNTTITVHYL-DLQLPFS 344
Qy 352 AP-ERQEGSSLTITCEASSODLEFQWLREBTGOVLERGVLQHLDKREAGGGRVAVS 410
Db 345 GEVTKQIEALPVSCITSSSRNATVFWIKDNTNR--MKTSP--SFSSLOYODAGNYICETT 400

Qy 411 VPSITGLNRTQLVNVAIFGPPWMAFKERKVVWKENVNLNLSCEASGHPRTISMNVNGTA 470
Db 401 HKEVEGLKKRRTKLIVGKRP--QIKMTKKTNTNKNKSKTIVCHVEGFPPKPAQMTVIGSG 458
Qy 471 S--EQDODPQVRLSTLN--VLVTPELLETVGECTASNDLGKNTLSLELVNLTLPDPS 526
Db 459 SLINKTEETKYVNGKFSKKIITIAPEENVTL-LNCIAENEL-----E 497
Qy 527 NNTTGLSTSTAS-----PHTRANSTSTERKLPPEPSRGVIVAVYICVLAVLAGAVLY 580
Db 498 RVTSLNVSATSIPEYDEPEDR--NDNSEK----VNDQAKLIVIGVGLLVALVAGVY 552
Qy 581 FLY-KKGKLLPCRRSGKQETILPPSRKSE 607
Db 553 WLTVKSKTSKXKHVDKLDGNIENKKE 580

RESULT 7
JH0464
DM-GRASP precursor - chicken
C:Species: Gallus gallus (chicken)
C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 02-Sep-2000
C:Accession: JH0464
R:Burns, F.R.; von Kaenen, S.; Guy, L.; Raper, J.A.; Kamholz, J.; Chang, S.
Neuron 7, 209-220, 1991
A:Title: DM-GRASP, a novel immunoglobulin superfamily axonal surface protein that sup
A:Reference number: JH0464; MUID:91337449
A:Accession: JH0464
A:Molecule type: mRNA
A:Residues: 1-587 <BUR>
A:Experimental source: brain
C:Comment: This is a cell surface glycoprotein.
C:Comment: This protein is localized to axons in the dorsal funiculus and ventral mid
C:Keywords: glycoprotein
F:1-32/Domain: signal sequence #status predicted <STO>
F:33-367/Product: DM-GRASP #status predicted <DMC>
F:67,198,270,311,365,461,484,503/Binding site: carbohydrate (asn) (covalent) #status

Query Match 14.9%; Score 502.5; DB 2; Length 587;
Best Local Similarity 26.3%; Pred. No. 2e-24;
Matches 165; Conservative 118; Mismatches 254; Indels 91; Gaps 30;
Qy 6 LVCAFLAACCCCPRVAGPEAEQAPAPELVEVEGSTALLKCGLSQSGNLSHVDW-FS 64
Db 17 LLC--LLALALCMPPALGL-----YTVNAVYGDITIMPCRLVDPG-LMFGKMYE 64
Qy 65 VHKEKRTLI-FR--VRQGGQSGEPGEYEOURLSLODRGATLALQVTPODERIFLCO--GK 119
Db 65 MPNGSPVFIAPFSSSTKKNQYDVPDKDRLSLSE--NTTSLKNARIRHKKRFVCMLYTE 123
Qy 120 RPRSGEYRIQLRVYKAAPEEPNI--QVNPGLIPVNSKEPEVATCYGRNGYPIPOVIWYKN 177
Db 124 DVVSEEPYV-KVYFKQSPQPEILHQADF-----ETEKLMGLGCVYRSDYPEGNVTWYKN 178
Qy 178 GRPLK--EKNRVHIQSSQTVESGLTYLQSLKQALVQKDAFQCELANRLPSGNHM 235
Db 179 GRVLPQVEEVVAVINLKRVEN-RSTGLFTMTSSLOYMPTEKDANAKFTCIYTVHGSGOKT 237
Qy 236 KESRETVVAVPYPEKRWLEV--EPVGMLEKGGDRVEIRCLADGNPPH--FSISKONPST 291
Db 238 IOSEPVEFVHYPTKEKTVIRVLSQSSSTIKEGNVTLKCGNGNPPPOEFLFYI----- 290
Qy 292 REAEETITNDNGVILVLPARKHSGRYECQGLDITMISLSEPOELLVNYSDVRVSPA 351
Db 291 -PGETEGIRSSDYVTMTVRNRNATGEYKCSLIDKSM-----DATITTVHYL-DLQLPFS 343
Qy 352 AP-ERQEGSSLTITCEASSODLEFQWLREBTGOVLERGVLQHLDKREAGGGRVAVS 410
Db 344 GEVTKQIEALPVSCITSSSRNATVFWIKDNTNR--MKTSP--SFSSLOYODAGNYICETT 399
Qy 411 VPSITGLNRTQLVNVAIFGPPWMAFKERKVVWKENVNLNLSCEASGHPRTISMNVNGTA 470

Db 400 LOEBEGLKKRTKLTIYEGRP--QIKMTKTNTNMKSKTIVCHEGSPKRAQVMTVSGS 457
Qy 471 S--EODODPOPVLTSLN--VIVPELLETGVECTASNDLGKNTSILELVNLTTLTPDS 526
Db 458 SLINKTEETKYVNGKFSKIIIAPEENVY--LTICIAENEL-----E 496
Qy 527 NMTTGLSTSTAS-----PHTRANSTERTKLPESRGVIVAVIICIIIVAGVAVLY 580
Db 497 RTVSLNVAISLIPEDYEDR--NDDNSEK-----VNDQAKLIVGIVGLLVALVAVY 551
Qy 581 FLY--KKGKLPGRSGKOEITLPPSRKSE 607
Db 552 WLIVYKSKTASAKHVDKDLGNIEENKMLE 579

RESULT 8
139428

alcam - human
C:Species: Homo sapiens (man)
C:Date: 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change 21-Jul-2000
C:Accession: 139428
R: Bowen, M.A.; Patel, D.D.; Li, X.; Modrell, B.; Malacko, A.R.; Wang, W.C.; Marguardt, H.
J. Exp. Med. 181, 2213-2220, 1995
A: Title: Cloning, mapping, and characterization of activated leukocyte-cell adhesion mol
A: Reference number: 139428; MUID:95279947
A: Accession: 139428
A: Status: preliminary; translated from GB/EMBL/DBJ
A: Molecule type: mRNA
A: Residues: 1-583 <RES>
A: Cross-references: GB:L38608; NID:9886257; PIDN:AAB59499.1; PID:9886258

Query Match 14.6%; Score 492.5; DB 2; Length 583;
Best Local Similarity 24.8%; Pred. No. 8.8e-24;
Matches 145; Conservative 121; Mismatches 236; Indels 63; Gaps 23;
Qy 41 GSTALKKGLSQSQGNLHVDFSVHKEKRTLIFRVRGOGSQ-----EPGEYEQRLSD 96
Db 36 GTTIIIPCLDVPQ--NLMFGRKKYKPKPGSPFIARFSTKSKSVGYDVPYKDRNLNSE 94
Qy 97 RGATLALQVTPQDERIFLCQ--GKRPSQOEYRIQLRYKKAPEEPNIOVNPGLIPVNSKEP 155
Db 95 -NYTLISNARISEDEKRFYCMVTEENYFEAPTIVKVPKOPSPKIVSKALFL--ETPOL 151
Qy 156 EEVATCVGNNGYPIQVIMYKNGRPLKEEKNRVHIQSQTVES--SGLYTLOSLKAOLVK 214
Db 152 KRLGDCISDSTPDGNTITWRNGKVLHLEGAVALIFKEMDPYQILTMTSTLEYKTK 211
Qy 215 EDKDAQFYCELNYRLPSGNHMKESREVTVPVPEYTEKYWLEV--EPVGMLEKGRVETRC 273
Db 212 ADIOWPFGSVTYGPGSKGTIHSQAVDPDIYPTQVLTIVLPKNAIKEDNTIKCL 271
Qy 274 AGGNPPH--FSISKQNSTREAEETINDNGVILPEARKHSGRGECGLDIDMTIS 330
Db 272 GKNPNPEPEFLYLPQPPGIRSSNTYTLMD-----VRNNAVGDYKCSLIDKKSMA 323
Qy 331 LNSEPELLVNVSDVRAVPAAP--EROEGSSLTITCEASSODLEFOMLREETGCVLERG 389
Db 324 STRA---ITVHTL--DLSLNPGEVTRIGDALPVSCITISASNNATVMMKDNIR--LWSS 376
Qy 390 PVLQHLDKREAGGYRCYASVSPIDGLNRKQLOLVNAITGPPWMAFEKKYVAKEMVNLN 449
Db 377 P--SFSSLHYODAGNVCEFTALQEVGLCKRRESLTILYVGP--QIKMTKTDPSGLSKT 432
Qy 450 LSCGASHGPRPTSMNVNTAS---EODODPO---RVLSLTVLVVPELLETGVECTASN 503
Db 433 IICHVAGFPKPAIQMTITSSGSYINOTESPIYNGRY--KIILSPENVT--LICTAEN 489
Qy 504 DLGKNTSILFELVNLTLTPDSNTTGLSTASPHTRANSTERTKLPESRGVAV 563
Db 490 QLRRTVNSLNVASISL---PE-----HDADEISDENR--EKVNDQAKLIV 530

Qy 564 AVIICVLIVAVGLVFLY--KKGKLPGRSGKOEITLPPSRKSE 607
Db 531 GIYVGLLALVAVGVYVWLMYMKSKTASAKHVKDGLGNNEENKMLE 575

RESULT 9
150478

neutrolin - goldfish (fragment)
C:Species: Carassius auratus (goldfish)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 13-Sep-1996
C:Accession: 150478
R: Jaesing, U.; Giordano, S.; Stecher, B.; Lottspeich, F.; Stuenkel, C.A.
Differentiation 56, 21-29, 1994
A: Title: Molecular characterization of fish neutrolin: a growth-associated cell surfac
protein DW-GRASP/SC-1/BEN.
A: Reference number: 150478; MUID:94299040
A: Accession: 150478
A: Status: preliminary; translated from GB/EMBL/DBJ
A: Molecule type: mRNA
A: Residues: 1-523 <LAEE>
A: Cross-references: GB:L25056; NID:9407318; PID:9407319

Query Match 13.6%; Score 457; DB 2; Length 523;
Best Local Similarity 26.7%; Pred. No. 1.4e-21;
Matches 158; Conservative 108; Mismatches 226; Indels 100; Gaps 25;

Qy 41 GSTALKK--GLSQSQGNLHVDFSVHKEKRTLIFRVRGOGQ-----SEPEYEQRLSL 94
Db 2 GETIVVPCNDGTKKRPG--LIFTWKYKVDGSGDLVLVQAKDETVSATGYSRVS 60
Qy 95 QDRGATLALQVTPQDERIFLCQ--GKRPSQOEYRIQLRYKKAPEEPNIOVNPGLIPVNSK 153
Db 61 A--ANSSLLIARGSLADQVFTCMVSEFTNLEEYSEVVKYKPSAVIK-----NNAK 112
Qy 154 EPE-----EVATCVGNNGYPIQVIMYKNGRPLKEEKNRVHIQSQTVES--SGLYTLOSL 207
Db 113 ELENGKLTQLGCEVNEANPADLIMKKNQTLVDGKTLITITLKTKITGLSTSSR 172
Qy 208 LKAOLVEDKDAQFYCELNYRLPSGNH-----KESREVTVPVPEYTEKYWLEVPEYGM 263
Db 173 LQYIARKEDEVESQTC-----TAKHVMPDQVSESPSPHYPTEKVSLOVVSQSP 225
Qy 264 EGDREYRCLADGNPP--HFSISKQNSTREAEETINDNGVILPEARKHSGRYC 320
Db 226 EGEDVTLKCADGNPPTSFNFNI-----GKKYTVDKQVYTLTGVTADSGIYC 277
Qy 321 QGLDLDMMISLSEPOBLVNVSDVRAVPAPE--ROEGSSLTITCEASSODLEFOMLR 379
Db 278 SLINDVME--TQEVTVSFL--DVSLLPTGKVLKAVGENLIVSDKNASSAKYTWK 332
Qy 380 EETGQVLERGFVLQHLDKREAGGYRCVAVSPIDGLNRTQLOLVNAITGPPWMA--FKE 437
Db 333 D--NRKLDKLP--DESKLTYSADAGLYCDV---SIEIKRSLSEFLVEGIPKITSLTRH 385
Qy 438 RKYVAKEMVNLNLCESAGHPRTISMNNGCTASEODQDQVRLSTNLVNVPELLETG 497
Db 386 RSSDGKRV--LICEAGSPKPDQVMSVNGCTNDEVSYNGKATYKLTVPVSNLT--V 439
Qy 498 ECTASNDLGKNTSILFELVNLTLTPDSNTTGLSTASPHTRANSTERTKLPES 557
Db 440 SCIVTNKIKGEDTKEL-----SVFSQKNKEDGTE----- 466
Qy 558 RGVIVAVICILVAVGLVFLY--KKGKLPGRSGKOEITLPPSRKSE 607
Db 467 QAKYIVGIVGLVLAALVGLIYWIYIKTROGSMKTGEAGTSEESKMLE 518

RESULT 10
120932

hypothetical protein F15G_4a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000

C:Accession: T20992; T24733
R:Sulston, J.
submitted to the EMBL Data Library, December 1994
A:Reference number: Z19355
A:Accession: T20992
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5175 <W11>
A:Cross-references: EMBL:Z47068; PIDN:CAA87335.1; GSPDB:GN00028; CESP:FL15G9.4a
A:Experimental source: clone F15G9
R:Kershaw, J.
submitted to the EMBL Data Library, December 1994
A:Reference number: Z19929
A:Accession: T24733
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5175 <W12>
A:Cross-references: EMBL:Z47070; PIDN:CAA87344.1; GSPDB:GN00028; CESP:FL15G9.4a
C:Gene: CESP:FL15G9.4a
A:Gene position: X
A:Introns: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184/3;
A:Exons: 2512/2; 2593/3; 2699/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 2991/3; 3033/1;
1; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4902/1; 4964/1; 5034/1; 51

Query Match 8.4%; Score 281; DB 2; Length 5175;
Best Local Similarity 22.1%; Pred. No. 5.6e-09;
Matches 122; Conservative 91; Mismatches 220; Indels 118; Gaps 27;

OY 27 EADOPAPELVEVVGSTAL-----LKGLSOSQGNLSHVMFVSKKERTLIFRVGQ- 79
Db 1357 DVGEPIILIPSTQNTTAVVGDVRLKCYVEASPP--ASYTWTF-----RRGI 1401
OY 80 -OGQSEPEYEQRLSLDRGATLALTYTPDDEIFLCQGRPRSQ-EYRQLRVYKAP 137
Db 1402 AIGTDTKG-----YVESDGLVIOASVEDATITTCASNPAKAEANLQVTVIASPD 1455
OY 138 --EPNIQVPLGIPVNSKPEEVATCVGRNGYPIPOVIMYKNGRPL-----KEKN 186
Db 1456 IKDPDVVTOE---SIKESHPFSLYCPVSN--PLQISWLYNDKPLIDDKTSWKTSDOKR 1510
OY 187 RVHIQSSQTVESGLTYTLOSILKAQVYKEDDAQFCYCELNRLPSGNHMKESR-EVTVPV 245
Db 1511 KLHVFKA-KITDSGVYKCAV-----RNAAGESKSKSQVEVITVPL 1548
OY 246 FYPTKEVMLEVEPVGMLEKGRDVEIRCLADGNPPPHFSI---SKONPSTREAEETTND 301
Db 1549 NLDSKYYKKV---FAKEGEVTLGCPVSGFPVPOIMWVVDGTVVEPGKKYKATLSND 1604
OY 302 NGVLYLEPARKHSGRYECOG-----LDLDTMISLSEPOELLVNYSDVRVSPAAPER 355
Db 1605 GLTLHFDSVSVKQEGNYHCVAQSKGNLIDIVELSVLAVP---IVGEDNLEVF----- 1655
OY 356 QEGSSLTITC--EASQDLEFOWL-----REETGOVLERGVLQDLHLKRAAGGGR 406
Db 1656 -LGDISTISCDLQTESDDKTTFFVWSINGESDRPDNVQIIPSDGRLITTAKPENNNKYM 1714
OY 407 CVASVPSTIGLNRQTLVNV--AIFGPPMAFKERKVVKENMVLNLSCEASGHPRTIS 463
Db 1715 CRVY-NSAGKAKERTLTDVLEPVPFVEP--VEFANQKLIGNNPII-LQCVTGNPKRTVI 1770
OY 464 WNVGTSSEODODOPQRYLSTLVNLTPELLETG---VECTASNDLGKNTSIFLELVNL 519
Db 1771 WKIDGNVDKSMLEDESLSLIRI---EKL-TGKSAQISCTAEKAKGTASRDFFIQIAA 1825
OY 520 TLTTPDSNTTT 530
Db 1826 PTFKNEGDQET 1836

RESULT 11

T43290
hemiscentin precursor - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000
C:Accession: T43290; T20993; T24734
R:Voegel, B.E.; Hedgecock, E.M.
submitted to the EMBL Data Library, June 1998
A:Description: Hemiscentin is required for hemidesmosome mediated cell adhesion and ge
A:Reference number: Z22396
A:Accession: T43290
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-5198 <VOG>
A:Cross-references: EMBL:AF074901; PIDN:AAC26792.1
R:Sulston, J.
submitted to the EMBL Data Library, December 1994
A:Reference number: Z19355
A:Accession: T20993
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5198 <W11>
A:Cross-references: EMBL:Z47068; PIDN:CAA87336.1; GSPDB:GN00028; CESP:FL15G9.4b
A:Experimental source: clone F15G9
R:Kershaw, J.
submitted to the EMBL Data Library, December 1994
A:Reference number: Z19929
A:Accession: T24734
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5198 <W12>
A:Cross-references: EMBL:Z47070; PIDN:CAA87345.1; GSPDB:GN00028; CESP:FL15G9.4b
A:Experimental source: clone T09B9
C:Gene: him-4; F15G9.4b
A:Map position: X
A:Introns: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184
A:Exons: 2512/2; 2593/3; 2699/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 2991/3; 303
1; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4902/1; 4964/1; 5034/1; 51

Query Match 8.4%; Score 281; DB 2; Length 5198;
Best Local Similarity 22.1%; Pred. No. 5.6e-09;
Matches 122; Conservative 91; Mismatches 220; Indels 118; Gaps 27;

OY 27 EADOPAPELVEVVGSTAL-----LKGLSOSQGNLSHVMFVSKKERTLIFRVGQ- 79
Db 1357 DVGEPIILIPSTQNTTAVVGDVRLKCYVEASPP--ASYTWTF-----RRGI 1401
OY 80 -OGQSEPEYEQRLSLDRGATLALTYTPDDEIFLCQGRPRSQ-EYRQLRVYKAP 137
Db 1402 AIGTDTKG-----YVESDGLVIOASVEDATITTCASNPAKAEANLQVTVIASPD 1455
OY 138 --EPNIQVPLGIPVNSKPEEVATCVGRNGYPIPOVIMYKNGRPL-----KEKN 186
Db 1456 IKDPDVVTOE---SIKESHPFSLYCPVSN--PLQISWLYNDKPLIDDKTSWKTSDOKR 1510
OY 187 RVHIQSSQTVESGLTYTLOSILKAQVYKEDDAQFCYCELNRLPSGNHMKESR-EVTVPV 245
Db 1511 KLHVFKA-KITDSGVYKCAV-----RNAAGESKSKSQVEVITVPL 1548
OY 246 FYPTKEVMLEVEPVGMLEKGRDVEIRCLADGNPPPHFSI---SKONPSTREAEETTND 301
Db 1549 NLDSKYYKKV---FAKEGEVTLGCPVSGFPVPOIMWVVDGTVVEPGKKYKATLSND 1604
OY 302 NGVLYLEPARKHSGRYECOG-----LDLDTMISLSEPOELLVNYSDVRVSPAAPER 355
Db 1605 GLTLHFDSVSVKQEGNYHCVAQSKGNLIDIVELSVLAVP---IVGEDNLEVF----- 1655
OY 356 QEGSSLTITC--EASQDLEFOWL-----REETGOVLERGVPVQLDLKRAAGGGR 406
Db 1656 -LGDISTISCDLQTESDDKTTFFVWSINGESDRPDNVQIIPSDGRLITTAKPENNNKYM 1714
OY 407 CVASVPSTIGLNRQTLVNV--AIFGPPMAFKERKVVKENMVLNLSCEASGHPRTIS 463

Db 1715 CRVT-NSAGKAEKTLTLDVLEPPVEEP--VEEANOGLIGNNPII-LOCOQTGNPKPLVI 1770
QY 464 MNVNSTASEODODPORVLTSTLVLTPELLEFG----VECTASNDLGNSTLLELVNL 519
Db 1771 WKIDONDYDKSMLEPDESLSLIRI-----EKL-TGSAQISCTAENKAGTASHDFIOMIAA 1825
QY 520 TTLPPDSNTTT 530
Db 1826 PTFKNEGDOET 1836

RESULT 12

A49448
Irregular chiasm C-roughnest precursor - fruit fly (Drosophila melanogaster)
N:Alternate names: irrec-roughnest protein
C:Species: Drosophila melanogaster
C>Date: 11-Nov-1994 #sequence_revision 11-Nov-1994 #text_change 07-May-1999
C:Accession: A49448; S34129
R:Ramos, R.G.P.; Igloi, G.L.; Lichte, B.; Baumann, U.; Meler, D.; Schneider, T.; Brandst
Genes Dev. 7, 2533-2547, 1993
A:Title: The irregular chiasm C-roughnest locus of Drosophila, which affects axonal proje
A:Reference number: A49448; MUID:94102535
A:Accession: A49448
A>Status: preliminary
A:Molecule type: DNA; mRNA
A:Residues: 1-764 <RAM>
A:Cross-references: GB:L11040; NID:g304790; PID:g304791; EMBL:Z21641; NID:g312985; PID:g
C:Genetics:
A:Gene: FlyBase:rst
A:Cross-references: FlyBase:FBgn0003285
C:Keywords: transmembrane protein

Query Match 8.0%; Score 270; DB 2; Length 764;
Best Local Similarity 20.7%; Pred. No. 2e-09;
Matches 137; Conservative 89; Mismatches 277; Indels 158; Gaps 25;

QY 40 VGSTALLKCGLSOSQGNLSHV-DMFSVHKERTLIIFRVROGQSGEPGEYFORLSIDRG 98
Db 41 VGARVTLPCRVYINKGQTLQMTKDFGLGTSRDLGSPFRVYAVNGSDDEGDY----- 90
QY 99 ATLATQVTPDDEKIFLCQK-----RPRSOEYRIQLRYKAPRPNIQVNPGLIPVNSK 153
Db 91 -SLDIYPMALDDARQCVQSPGEPFALRSTFAGLTVLVPAPAPKITQGDVIYATADR 149
QY 154 EPEEVATCGVNGYPIPOVLYYKN-GR-----PLKEKNRHHQSSQVVESSGL 201
Db 150 KVE--TECVSVGKRAALITWIDGLGNVLDNIETVYIPLPDOR----- 192
QY 202 YTLQSLKAQLVKEDKAQFYCELNRYLRPSGNHMKESREVTVPVFPTEKVMLEVEPVGM 261
Db 193 FTASVSLTLPRKHHNTNFSCQAQ--NTADRTYRSKAIKEVKYAP--YAVNNMGS 246
QY 262 LKEG-----DRVEINCLADGNPP--HPSISKONPSTREAE 295
Db 247 LPGAAGSVGAGAGSVHMTSGSRIVESHQVRLRCRADANSDVRYWFIINDEPIIGOK 306
QY 296 EETTNDNGVLVLEPARKHSGRQCGGLDITMISLSEPELLVNVYSDVRSPAPAPR 355
Db 307 TE-----MVRNVTYRKHDALIVK--EVQNSVGSSEDSSTLDTISTAPSRQRPQSEMA 357
QY 356 QEGSSLTLTCEASSODLEFQWLREETGOVLERGPVLQDLHLKREAGGVCVAVSPESIP 415
Db 358 DVGSVSLTCEVDSNPQEIYWIQHPSDRVYGTSTNLTF--SVSNETAGRYTCRKNVGYA 416
QY 416 GLNRQLVNVALIFGPPMAFEKRVVWKENVYLMSCASGHPPT--ISNVNG--TASE 472
Db 417 EISADAVYLLK--GSPAIG--SQRTQYGLVGDITARIECFASVPRARHVSWTFNGQETISE 473
QY 473 QDOD-----PQVLTSTLVNLTPELLETGECTASNDLGNSTLLELVNLTLTLP 524
Db 474 SGHDYSLIYDAVPGGVKSTLIIRDSQAYHIGKYNCYVNDIGNDVAETLOAKKSVSL-- 531

QY 525 DSNWTTGLSTSTASPHTRANSTSTERKLEPESRGVYIVAVIYCIJVLAVLAEFLYK 584
Db 532 -----LMTIVG-----ISVAFLLVLTIL-VVY----- 555
QY 585 KGLPCRRSGKQETLPPSRKSELVVEYKSKLPEEMGLQSSGDKRAPGDQGERYIDL 644
Db 556 ---IKCKRRKTLPP-----ADVISEHQITKNGVSKLEPGRGRTSNYSDL 597
QY 645 R 645
Db 598 K 598

RESULT 13

JC2457
Vascular cell adhesion protein - pig
C:Species: Sus scrofa domestica (domestic pig)
C>Date: 15-Feb-1995 #sequence_revision 05-Apr-1995 #text_change 08-Oct-1999
C:Accession: JC2457
R:Tsang, Y.T.M.; Haskard, D.O.; Robinson, M.K.
Biochem. Biophys. Res. Commun. 201, 805-812, 1994
A:Title: Cloning and expression kinetics of porcine vascular cell adhesion molecule.
A:Reference number: JC2457; MUID:94271236
A:Accession: JC2457
A:Molecule type: mRNA
A:Residues: 1-538 <TSA>
A:Cross-references: EMBL:008351; NID:g474382; PID:AAA21542.1; PID:g474383
C:Keywords: glycoprotein; transmembrane protein
F:497-517/Domain: transmembrane status predicted <TM>
F:75,157,271,330,360/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 7.9%; Score 264.5; DB 2; Length 538;
Best Local Similarity 21.6%; Pred. No. 2.8e-09;
Matches 166; Conservative 103; Mismatches 234; Indels 119; Gaps 27;

QY 31 PABEIVEVEGSTALLKCGLSOSQGNLSHVDMFSVHKERTLIIFRVROGQSGEPGEYEQ 90
Db 31 PEKMI-AQIGDASLSLTCAPDCESSLS-FSW-----RTQIDSPNG----- 70
QY 91 RLSLDQRGATLALQVTPDDEKIFLC-----QGRPRSOEYRIQLRYKAPRPNIQV 143
Db 71 KVTNGTRSTLVNPNVSPFENEHSLCTVSCGNLKGR-----GIQVETISPPKPEIHW 124
QY 144 NPL---GIPIVNSKEPEVATCYGRNGYPIR--QYIYKNGRPLKBEK--NRHHQSSQTV 196
Db 125 SSLPEVAKPYVAR-----CLVPDVPVEKLEIELLNKNSHMSVQNFLELIDISKETK 177
QY 197 ESSGLYTLQSLKAQLVKEDKAQFYCELNRYLRPSGNHMKESREVTVPVFPTEKVMLEV 256
Db 178 SLEFTT-----PTFEDIGALYCAQATLIIIDQSPSKYTPPEKMYIISKDVP-ISV 228
QY 257 EYVGMLEKGDVREIRCLADGNPPRPHRSISKONPSTREAEETTNDNGVLVLEPARKHSG 316
Db 229 NPSTLSQEGDSMMATCTSGELPAPQISMGK--LDNGDQQLLSGATLTLTAMMEDSG 285
QY 317 RYECQGLDITMISLSEPELLVNVY---SDVRSPAPRQEGSSSLTTCSEASSODL 373
Db 286 IYCEGVN--PVGNRKEVELTVQAPDQTTISVNPSS-TLEEGSVMTGSSSGFPAP 341
QY 374 EFQW---LREETGOVLERGPVLQDLHLKREAGGVCVAVSPISGLNTOVLVNAIFGP 430
Db 342 KILMSKRLRGNLEPSEMTTLITSTKNEDSGIYCEBI--NAGINNKE-VELIQA 398
QY 431 P---WMAFERKRVVWKENVYLMSCASGHPPTISMNWNGTASQDODPORVLTSTLV 486
Db 399 PKDQLTAPPSVESYKKGDIIVISCTC--GNVPPTLI-----ILKKAETGDTVLKSTDG 450
QY 487 LVT---PELLEGV-ECTASNDLGNSTLLELVNLTLTLPDSNWTGSLSTASPHTR 542
Db 451 AYTIHRRALADAGVYECSEKNEIG-----LQLRSTLTDVKGRESN----- 490

QY 543 ANSTERKLPESRGVIVAVICILAVLAVLYELK 584
Db 491 KDYFSE-----LVLYCASSLIIRAIIVITFAK 521

RESULT 14

A39712
kinase-like protein k1g precursor - chicken
C:Species: Gallus gallus (chicken)
C:Date: 08-Nov-1991 #sequence_revision 08-Nov-1991 #text_change 24-Sep-1999
C:Accession: A39712
R:Chou, Y.H.; Hayman, M.J.
Proc. Natl. Acad. Sci. U.S.A. 88, 4897-4901, 1991
A:Title: Characterization of a member of the immunoglobulin gene superfamily that possib
A:Reference number: A39712; MUID:91271300
A:Accession: A39712
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1051 <CHO>
A:Cross-references: GB:M63437; NID:9212235; PIDN:AAA4893.1; PID:9212236
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo
C:Keywords: ATP
E:775-1046/Domain: protein kinase homology <KIN>
E:783-791/Region: protein kinase ATP-binding motif

Query Match 7.7%; Score 259.5; DB 2; Length 1051;
Best Local Similarity 20.1%; Pred. No. 1.5e-08;
Matches 149; Conservative 89; Mismatches 26; Indels 241; Gaps 31;

QY 29 EQPABELVEVGSSTALIKGLISQSGNLSHVD-----WF-----SVHK 67
Db 119 KQPA-SAAEIPGSSITVILRC-----HIDGHPRTWQFRCGAPLPGRGYSVS 167
QY 68 EKRTILFRVROQSGSEGEY-----EQRLSDRGATLALTQ-----Y 106
Db 166 KERTTLR---GAGPDDNCLLYCSARPRAVGVCSDNFTNIIIDESPQAVVPEDLI 224
QY 107 TQODERIFLCQ--GKRPRSOEY-----RIQLR---YK--- 134
Db 225 TNEEMAFPCQPAVPRPQDELFDSPITNRSKTTVANGSLTLQYKASTGYKKGIG 284
QY 135 -----APEEPNIQVNPICIPYNSKEPEEVATCGVGRNGYPIQV 172
Db 285 HGQSKALVLTAKTLRLAEIEEMAPSPVLTANQGHV-----SCACPRGVPTQV 335
QY 173 IYKKN-----GRPLKEKNRVHIQSSQTVSSGLTYLQSLTKAQVKEKDAQFYCEL 225
Db 336 IWERKQERVPYTAGRVYQEAQLV---TSTTEADAGITVCHANKA----- 378
QY 226 NVRLPSGNHMKESREVTVEFYPTKEVMLEVPRVGMLEKGEVETRCLADGNPPHFSIS 285
Db 379 -----GKKQELSTV---ATVPKVMEMKXDSLESKRGLYHCHSKASLKTIVMY 427
QY 286 KNPSTREAEETITNDGNVLEPARKEHSGRYECOGDLDTMISLSEPOELLVNY--- 342
Db 428 RNVGSISSEDSREISENGTLRIINVEYVDGTWYKC-----VSTPAGSIGIYARV 477
QY 343 ---VSQVRVAPARERQ---EGSSLTLCFAESSQDLEFQWLMREE---TGQVLESGPIQ 393
Db 478 HYLEKTKKFTPPPOPLQCMENKEVTVSCATGREKPTIOWTKTDSSLSHVSAGILIS 537
QY 394 LHDLKREAGGRCYAVASVSIPLNRTOLVNAIFGPPMAFK--ERKVVYKENVNLIS 451
Db 538 FIKVSSDSGNGTCTIASSNPOGEIRATVQLVAVY---VFKEKLEPPTVYQGHATMFQ 593
QY 452 CEASGHPRTISMNVNGTASEQDOPQRYLSTLVNLVLPPELLETVECTASDLCKNTSI 511
Db 594 COAEGDPVHIQWK---GKDKILDPSSKILPRIQIIMPNGSLV---IYDVTEDSGKYTCI 646
QY 512 -----LLELVNLTITLPDSNTTGTSTASPHRTANSTERKLPPEESGV 560
Db 647 AGNSCNIKHREAFLLVYVDPAAEEDEG-----PSSHPTYKMIQT-----IGL 688

QY 561 VIVAVICILAVLAVLYELKKGKLPGRSGKOETLPPRSKSEL----- 608
Db 689 SVGAAYVITII-VLG--LMFYCK-----RRAKRLKHPREGEPEMECLNGTILLONG 739

QY 609 --VVEKSKDLPEEMGLIGSS 628

Db 740 QTVAIEQEEVALNMLGSSSGAS 761

RESULT 15

IJXNL
neural cell adhesion molecule long domain form precursor - African clawed frog
N:Alternate names: NCAM-180
N:Contents: neural cell adhesion molecule, short domain form (NCAM-140)
C:Species: Xenopus laevis (African clawed frog)
C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 22-Jun-1999
C:Accession: S09600
R:Krieg, P.A.; Sakaguchi, D.S.; Kintner, C.R.
Nucleic Acids Res. 17, 10321-10335, 1989
A:Title: Primary structure and developmental expression of a large cytoplasmic domain
A:Reference number: S09600; MUID:90098871
A:Accession: S09600
A:Molecule type: mRNA
A:Residues: 1-1088 <KRI>
A:Cross-references: EMBL:M25696; NID:9214609; PIDN:AAA49909.1; PID:9214610
A:Note: the authors translated the codon AAA for residue 970 as Leu
C:Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM
C:Genetics: Several forms of NCAM are produced by alternative splicing.

A:Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; 1
C:Keywords: alternative splicing; brain; cell adhesion; duplication; heparin binding;
E:1-19/Domain: signal sequence #status predicted <SIG>
E:20-1088/Product: neural cell adhesion molecule, long domain form #status predicted
E:20-803,1050-1088/Product: neural cell adhesion molecule, short domain form #status
E:34-95/Domain: extracellular #status predicted <EXT>
E:129-188/Domain: immunoglobulin homology <IMM1>
E:149-153/Region: heparin binding #status predicted
E:158-162/Region: heparin binding #status predicted
E:225-284/Domain: immunoglobulin homology <IMM3>
E:317-381/Domain: immunoglobulin homology <IMM4>
E:413-475/Domain: immunoglobulin homology <IMM5>
E:512-589/Domain: fibronectin type III repeat homology <FN3A>
E:618-679/Domain: fibronectin type III repeat homology <FN3B>
E:706-723/Domain: transmembrane #status predicted <TM>
E:724-1088/Domain: intracellular #status predicted <INT>
E:41-93,136-186,232-282,323-379,420-473/Disulfide bonds: #status predicted
E:219,310,341,417,443,472/Binding site: carbohydrate (Asn) (covalent) #status predict

Query Match 7.6%; Score 256; DB 1; Length 1088;
Best Local Similarity 18.1%; Pred. No. 2.6e-08;
Matches 155; Conservative 100; Mismatches 26; Indels 336; Gaps 35;

QY 33 PELVEVEGSTALIKGLISQSGNLSHVDVSVHKEKRTILFRVROQSGSEGEYQRL 92
Db 26 PDQGEISLGESEKFLC---QVSGEATDISWYSPTEKELV-----QQQL 66
QY 93 SL---QDRGATLALTQVTPQDERIFLC-----QK-----RPSOEY 126
Db 67 SVVSSDYTSITLTLYNASODAGIYKCVASNEAGESEGTVMKIKYOKLTFKNAPTPOE 126
QY 127 R-----IOLRYKAEDE----- 138
Db 127 KEGEDAVIICDVSSISITTRHKGDVTFKKDVRPVVANNYLQIRGKTKKTDDEGTYRC 186
QY 139 -----PNIQVNPICIPYNSKEPEEVATCGVGRNGYPIQVYIWK 176
Db 187 EGRILARGEINKDIQVIVVNPPTIQAQLRVNATAMASVVLSCDADGFPDPETISLWK 246
QY 177 NGRPLKEKNRVHIQSSQTVSSGLTYLQSLTKAQVKEKDAQFYCELNTRKLPSSGNHK 236

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2002, 10:46:35 ; Search time 14.84 Seconds

(without alignments)
1585.501 Million cell updates/sec

Title: US-09-653-961-2

Perfect score: 3363

Sequence: 1 MGLPRLVCAFLAACCPCPR.....SSGDKRAPGDGEEKYIDLRH 646

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3327	98.9	646	1	MU18_HUMAN
2	647.5	19.3	628	1	LU_HUMAN
3	509.5	15.2	588	1	C166_CHICK
4	492.5	14.6	583	1	C166_HUMAN
5	488.5	14.5	583	1	C166_MOUSE
6	465.5	13.8	555	1	C166_CARAU
7	438	13.0	564	1	C166_BRARE
8	270	8.0	764	1	ICCR_DROME
9	259.5	7.7	1051	1	PKK7_CHICK
10	256	7.6	1088	1	NCA1_XENLA
11	246.5	7.3	725	1	NCA2_MOUSE
12	246.5	7.3	739	1	NCA1_RAT
13	244.5	7.3	858	1	NCA1_RAT
14	244.5	7.3	3707	1	PGBM_MOUSE
15	242.5	7.2	853	1	NCA1_BOVIN
16	237	7.0	1092	1	NCA2_XENLA
17	235.5	7.0	1091	1	PGBM_HUMAN
18	235	7.0	1093	1	NCA1_HUMAN
19	234	7.0	761	1	NCA2_HUMAN
20	234	7.0	848	1	NCA1_HUMAN
21	233.5	6.9	1115	1	NCA1_MOUSE
22	232.5	6.9	1260	1	NCA1_MOUSE
23	230	6.8	1260	1	NCA1_MOUSE
24	230	6.8	739	1	VCA1_HUMAN
25	228	6.8	837	1	NCA2_MOUSE
26	228	6.8	873	1	FS21_DROME
27	226.5	6.7	1259	1	CAML_RAT
28	226	6.7	738	1	PECL_HUMAN
29	225.5	6.7	1257	1	CAML_HUMAN
30	225	6.7	702	1	CEAS_HUMAN
31	224.5	6.7	1070	1	PKK7_HUMAN
32	218.5	6.5	404	1	RAGE_HUMAN
33	217.5	6.5	1709	1	SN_HUMAN

34	215	6.4	1284	1	NRCA_CHICK	P35331	gallus gall
35	214	6.4	847	1	CD22_HUMAN	P20273	homo sapien
36	210	6.2	1333	1	VGR1_MOUSE	P35969	mus musculus
37	209	6.2	862	1	CD22_MOUSE	P35329	mus musculus
38	208	6.2	739	1	VCA1_MOUSE	P29533	mus musculus
39	207	6.2	526	1	CEAL_HUMAN	P13688	homo sapien
40	205.5	6.1	402	1	RAGE_RAT	P63495	rattus norv
41	205.5	6.1	1356	1	VGR2_HUMAN	P35968	homo sapien
42	204.5	6.1	416	1	RAGE_BOVIN	Q28173	bos taurus
43	204.5	6.1	2012	1	NSCA_HUMAN	O60469	homo sapien
44	203.5	6.1	1266	1	NGCA_CHICK	Q03696	gallus gall
45	203	6.0	1447	1	DCC_MOUSE	P70211	mus musculus

ALIGNMENTS

RESULT	ID	STANDARD	PRT	646 AA.
1	MU18_HUMAN			
1	PA3121:			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	01-MAR-2002 (Rel. 41, Last annotation update)			
DE	Cell surface glycoprotein MUC18 precursor (Melanoma-associated antigen MUC18) (Melanoma-associated antigen A32) (S-endo 1 endothelial-associated antigen) (CD146 antigen) (Melanoma adhesion molecule).			
DE	MCAM OR MUC18.			
GN	Homo sapiens (Human).			
OS	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RP	TISSUE=Melanoma;			
RX	MEDLINE=9009368; PubMed=2602381;			
RA	Lehmann J.M., Riettmueller G., Johnson J.P.;			
RT	"MUC18, a marker of tumor progression in human melanoma, shows sequence similarity to the neutral cell adhesion molecules of the immunoglobulin superfamily.";			
RT	Proc. Natl. Acad. Sci. U.S.A. 86:9891-9895(1989).			
RL	[2]			
RN	REVISIONS, SEQUENCE FROM N.A.			
RP	TISSUE=Melanoma;			
RX	MEDLINE=93391384; PubMed=8378324;			
RA	Sers C., Kirsch K., Rothbacher U., Riettmueller G., Johnson J.P.;			
RT	"Genomic organization of the melanoma-associated glycoprotein MUC18: implications for the evolution of the immunoglobulin domains.";			
RT	Proc. Natl. Acad. Sci. U.S.A. 90:8514-8518(1993).			
RL	[3]			
RN	SEQUENCE OF 24-44; 98-112; 135-153; 240-260; 379-389 AND 460-478.			
RP	MEDLINE=9415196; PubMed=8162602;			
RX	Shih I.-M., Elleder D.E., Speicher D., Johnson J.P., Herlyn M.;			
RA	"Isolation and functional characterization of the A32 melanoma-associated antigen.";			
RT	Cancer Res. 54:2514-2520(1994).			
RL	[4]			
RN	SEQUENCE OF 27-40; 98-112 AND 236-260.			
RP	MEDLINE=96136302; PubMed=8573133;			
RX	Bardin N., Frances V., Lesaule G., Horschowski N., George F.,			
RA	Sampl R.;			
RT	"Identification of the S-Endo 1 endothelial-associated antigen.";			
RT	Biochem. Biophys. Res. Commun. 218:210-216(1996).			
RL	[5]			
RN	FUNCTION.			
RP	MEDLINE=94122526; PubMed=8292890;			
RX	Johnson J.P., Rothbacher U., Sers C.;			
RA	"The progression associated antigen MUC18: a unique member of the immunoglobulin supergene family.";			
RT	Melanoma Res. 3:337-340(1993).			
RL	CELLS DURING EMBRYONIC DEVELOPMENT. ITS EXPRESSION MAY ALLOW			
CC	MELANOMA CELLS TO INTERACT WITH CELLULAR ELEMENTS OF THE VASCULAR			

SYSTEM THEREBY ENHANCING HEMATOGENOUS TUMOR SPREAD.
 CC SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: MAY APPEAR AT THE SURFACE OF NEURAL CREST
 CC CELLS DURING THEIR EMBRYONIC MIGRATION. APPEARS TO BE LIMITED TO
 CC VASCULAR SMOOTH MUSCLE IN NORMAL ADULT TISSUES. ASSOCIATED WITH
 CC TUMOR PROGRESSION AND THE DEVELOPMENT OF METASTASIS IN HUMAN
 CC MALIGNANT MELANOMA. EXPRESSED MOST STRONGLY ON METASTATIC LESIONS
 CC AND ADVANCED PRIMARY TUMORS AND IS ONLY RARELY DETECTED IN BENIGN
 CC MELANOCYTIC NEVI AND THIN PRIMARY MELANOMAS WITH A LOW PROBABILITY
 CC OF METASTASIS.
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
 CC -1- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE V-TYPE DOMAINS.
 CC -1- DATABASE: NAME-PROW; NOTE-CD guide CD146 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd146.htm"
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M29277; AAA20824.1; -;
 DR EMBL: M28882; AAA20922.1; -;
 DR EMBL: X68264; CAA48332.1; -;
 DR EMBL: X68265; CAA48332.1; JOINED.
 DR EMBL: X68266; CAA48332.1; JOINED.
 DR EMBL: X68267; CAA48332.1; JOINED.
 DR EMBL: X68268; CAA48332.1; JOINED.
 DR EMBL: X68270; CAA48332.1; JOINED.
 DR EMBL: X68271; CAA48332.1; JOINED.
 DR MIM: 155735; -;
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003598; Ig_C2.
 DR InterPro: IPR003600; Ig_Like.
 DR Pfam: PF00047; Ig; 5.
 DR SMART: SM00410; Ig_Like; 2.
 DR SMART: SM00408; IGC2; 2.
 KW Cell adhesion; Immunoglobulin domain; Glycoprotein; Transmembrane;
 KW Repeat; Signal;
 FT SIGNAL 1 23
 FT CHAIN 24 646
 FT DOMAIN 24 559
 FT TRANSMEM 560 583
 FT DOMAIN 584 646
 FT DOMAIN 39 122
 FT DOMAIN 154 229
 FT DOMAIN 265 327
 FT DOMAIN 358 414
 FT DOMAIN 445 506
 FT DISULFID 48 116
 FT DISULFID 161 223
 FT DISULFID 272 320
 FT DISULFID 365 407
 FT DISULFID 452 499
 FT CARBOHYD 56 56
 FT CARBOHYD 418 418
 FT CARBOHYD 449 449
 FT CARBOHYD 467 467
 FT CARBOHYD 508 508
 FT CARBOHYD 518 518
 FT CARBOHYD 527 527
 FT CARBOHYD 544 544
 SQ SEQUENCE 646 AA; 71793 MW; F064A5DAEB0BADEC6 CMC64;

Query Match 98.9%; Score 3327; DB 1; Length 646;
 Best Local Similarity 98.9%; Pred. No. 3.4e-212;
 Matches 639; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 MGLPLVCAFLIAACCCPRVAGVGEAEQAPAEVVEVSTALLKGLSGSOGNSLHV 60

|||||
 Db 1 MGLPLVCAFLIAACCCPRVAGVGEAEQAPAEVVEVSTALLKGLSGSOGNSLHV 60
 OY 61 DMFSVAKERTLIFRVQGGSEGEVEQRLSLDRATATLQVYQDRIFLFCQGR 120
 |||||
 Db 61 DMFSVAKERTLIFRVQGGSEGEVEQRLSLDRATATLQVYQDRIFLFCQGR 120
 OY 121 PPSQEVRIQRYKAPPEEPIQVNFGLIPVNSKEPEEVAATGVRNGYPIQVYWKNGR 180
 |||||
 Db 121 PPSQEVRIQRYKAPPEEPIQVNFGLIPVNSKEPEEVAATGVRNGYPIQVYWKNGR 180
 OY 181 LEEKNRVHIQSSQYVESSGLTYTLOSILKAOLVKEDKAQFYCELYNLPSGNHKKESRE 240
 |||||
 Db 181 LEEKNRVHIQSSQYVESSGLTYTLOSILKAOLVKEDKAQFYCELYNLPSGNHKKESRE 240
 OY 241 VIVPVYPTREKWLVEPEVGMLEKEDRVEIRCLADGNPPHSISKONPSTREAEETTN 300
 |||||
 Db 241 VIVPVYPTREKWLVEPEVGMLEKEDRVEIRCLADGNPPHSISKONPSTREAEETTN 300
 OY 301 DNGVLVEPARKHSGRYECQGLDPTMTLSLSEPOELLVNVSDVRSVPAAPERQEGSS 360
 |||||
 Db 301 DNGVLVEPARKHSGRYECQGLDPTMTLSLSEPOELLVNVSDVRSVPAAPERQEGSS 360
 OY 361 LRTTCAESSQDLFEQWLEEFQVLERGPVQLHDKREACGGRVCAVSPSTIGLNRT 420
 |||||
 Db 361 LRTTCAESSQDLFEQWLEEFQVLERGPVQLHDKREACGGRVCAVSPSTIGLNRT 420
 OY 421 QLVNVAIFGPWMAFKERYVWKENVNLSCASGHPRTISMVNGTASBODDPORV 480
 |||||
 Db 421 QLVNVAIFGPWMAFKERYVWKENVNLSCASGHPRTISMVNGTASBODDPORV 480
 OY 481 LSTLVNLVPELLETGVECTASNDLKNNTIFLELVNLTITLTPSNTTGTSTASGH 540
 |||||
 Db 481 LSTLVNLVPELLETGVECTASNDLKNNTIFLELVNLTITLTPSNTTGTSTASGH 540
 OY 541 TRANSTSTERKLPESRGRVYAVYVCLVAVYLYFYKKGKLPGRSGKQETL 600
 |||||
 Db 541 TRANSTSTERKLPESRGRVYAVYVCLVAVYLYFYKKGKLPGRSGKQETL 600
 OY 601 PPSRSELVEVKSDKLPPEEMGLQSSGDKRAPODGEKYIDLRH 646
 |||||
 Db 601 PPSRSELVEVKSDKLPPEEMGLQSSGDKRAPODGEKYIDLRH 646
 RESULT 2
 LU_HUMAN STANDARD; PRT; 628 AA.
 ID ID_HUMAN
 AC P50895;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Lutheran blood group glycoprotein precursor (B-CAM cell surface
 DE glycoprotein) (Aubberger B antigen) (F8/G253 antigen).
 GN LU OR BCM OR MSK19.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 32-67 AND 182-203.
 RC TISSUE=Placenta;
 RX MEDLINE=95296337; PubMed=7777537;
 RA Parsons S.F., Mallinson G., Holmes C.H., Houlihan J.M., Simpson K.L.,
 RA Mawby W.J., Spurr N.K., Warne D., Barclay A.N., Antee D.J.,
 RA "The Lutheran blood group glycoprotein, another member of the
 RT immunoglobulin superfamily, is widely expressed in human tissues and
 RT is developmentally regulated in human liver".
 RL Proc. Natl. Acad. Sci. U.S.A. 92:5496-5500(1995).
 [2]
 RN SEQUENCE OF 1-588 FROM N.A.
 RP MEDLINE=95042297; PubMed=7954395;
 RX Campbell I.G., Foulkes W.D., Senger G., Trowsdale J.,
 RA Garin-Chesa P., Rettig W.J.;

DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE CD166 antigen precursor (activated leukocyte-cell adhesion molecule)
 DE ALCAM (DM-GRASP protein).
 GN ALCAM.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCB1_TaxId=10090.
 RN NCB1
 RP SEQUENCE FROM N.A.
 RC STRAIN=NF5;
 RX MEDLINE=97353242; PubMed=9209500;
 RA Bowen M.A., Bajorath J., D'Egidio M., Whitney G.S., Palmer D.,
 RA Kobarg J., Starling G.C., Stidak A.W., Aruffo A.;
 RT "Characterization of mouse ALCAM (CD166): the CD6 binding domain is
 RT conserved in different homologs and mediates cross-species binding";
 RL Eur. J. Immunol. 27:1469-1478(1997).
 RN [2]
 RP SEQUENCE OF 227-583 FROM N.A.
 RC STRAIN=BALB/C; TISSUE=Brain;
 RX MEDLINE=94376084; PubMed=8089660;
 RA Kanli J.P., Chang S., Kuwada J.Y.;
 RT "The molecular cloning and characterization of potential chick
 RT J. Neurobiol. 25:831-845(1994).
 RL J. Neurobiol. 25:831-845(1994).
 CC -1- FUNCTION: CELL ADHESION MOLECULE THAT BINDS TO CD6. INVOLVED IN
 CC NEURITE EXTENSION BY NEURONS VIA HETEROPHILIC
 CC INTERACTIONS. MAY PLAY A ROLE IN THE BINDING OF T AND B CELLS TO
 CC ACTIVATED LEUKOCYTES, AS WELL AS IN INTERACTIONS BETWEEN CELLS OF
 CC THE NERVOUS SYSTEM.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
 CC -1- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE V-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE V-TYPE DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: U95030; AAC06342.1; -;
 DR EMBL: L25274; AAA37528.1; -;
 DR HSSP: Q13740; IRLC.
 DR MGD: MGI:131326; ALCAM.
 DR InterPro: IPR003599; Iq.
 DR InterPro: IPR003006; Iq_MHC.
 DR InterPro: IPR003600; Iq_Like.
 DR Pfam: PF00047; Iq; 5.
 DR SMART: SM00409; Iq; 3.
 DR SMART: SM00410; Iq_Like; 2.
 DR PROSITE: PS00290; Iq_MHC; FALSE_NEG.
 KW Cell adhesion; Immunoglobulin domain; glycoprotein; Transmembrane;
 KW Repeat; Signal.
 FT SIGNAL 1 27 POTENTIAL.
 FT CHAIN 28 583 CD166 ANTIGEN.
 FT DOMAIN 28 527 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 528 529 POTENTIAL.
 FT DOMAIN 530 549 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 549 583 IG-LIKE C2-TYPE DOMAIN 1.
 FT DOMAIN 583 607 IG-LIKE C2-TYPE DOMAIN 2.
 FT DOMAIN 607 633 IG-LIKE C2-TYPE DOMAIN 3.
 FT DISULFID 43 113 POTENTIAL.
 FT DISULFID 157 220 POTENTIAL.
 FT DISULFID 270 313 POTENTIAL.
 FT DISULFID 354 392 POTENTIAL.
 FT DISULFID 435 485 POTENTIAL.
 FT CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 167 167 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 265 265 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 306 306 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 361 361 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 480 480 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 227 232 PSQKRT -> AAGIPA (IN REF. 2).
 FT CONFLICT 454 454 S -> F (IN REF. 2).
 SO SEQUENCE 583 AA; 65161 MW; E7BAFA8FCA8F9489 CRC64;

 Query Match 14.5%; Score 486.5; DB 1; Length 583;
 Best Local Similarity 24.5%; Pred. No. 5,9e-25;
 Matches 152; Conservative 122; Mismatches 274; Indels 73; Gaps 25;

 QY 5 RLVCAPFLAACCCCPKAGVGEAEOPAPRLVVEVGSATLKGISQSGNLSDHDPFS 64
 DB 10 RLVPCLLSAVALRPLGLWV-----TVSAVGDTIVMCRDLVQ-NLMEFKMY 58
 QY 65 VHKRRTLIFRVGQGSQGS---EPEYERLSLDGATLQVTPQDERITLQ-GK 119
 DB 59 EKPDSVVFIAFRSSFTKSYQYDDEYKRLSLSE-NYTLISANKISDEKRFVCLVY 117
 QY 120 RPSQERIDLVYKAPKEENIOVNPGLIPVNSKEPEVATCGRGYPIPVYWKNGR 179
 DB 118 EDNVEAPTLVYKVPKQSPKEI-VNKAAP-LETDQLKLGDCISDSYDPGNTWYRNEK 175
 QY 180 PLKEKRRVYIQQSQVTE-SSGLYTLQSLKAQVLEDKDAQFYCLNRLPSGNMKES 238
 DB 176 VLQVBEVAILRKKEIDPQTLVYSSLEYTTSDIDMPGTCVTVYGPQGTIIS 235
 QY 239 REVTVPEVPTKQWLEV-EPVGMKRGDREYIRCLADGNPPH---FSTKONPSTREA 294
 DB 236 EQEIFDIYPTQEVTLQVLPKNAIKEDGNITLQCLGNGPPEEFMFYLPQGPESIRSS 295
 QY 295 EERTNNGVYVLEPAKSHSGRVEGQGLDLMISLSRQGLVNVYSDVSPAAV- 353
 DB 296 NYTYLTD-----VARNATGDKCSLIDKRNMAASTT---ITYVYL-DLSLNREGV 342
 QY 354 ERQGSLLTTCENESSQDLFEQMLRETEQVLERGPVQLDHLKREAGQVRCVAVSPS 413
 DB 343 TKQIGDLPVPSCTISARNATVYMKDNIR--LRSSP--SFSSLYHQDAGNYCETALDE 398
 QY 414 IPGINRQVLYVNAIFCPMAFKERKAYKENVNLSCASGHPRTISMVNGIAS-- 471
 DB 399 VEGIKKRESLTLVEGKP-QIKMTKTDPSGSKTIIICVEGPPRAIHWITTGGSVY 456
 QY 472 -EODDPQ--RYLSTLVNLTPELLETGVECTASNDIGKNTSILFLELVNLTLPDSDN 527
 DB 457 NQTESPYNGRIYS--KIISPEENVY-LTCTAENDLETVNSLVNSAIST---PE-- 507
 QY 528 TTGGLSTSTASPHTRANSSTERKLPEPESRGVYVAVICILVAVLYELV-KRG 586
 DB 508 -----HDEADDISDENR-EKVNDQAKIYGIYVGLLAVAGVYVLYMKRS 554
 QY 587 KLPRBSGQETIIPSRKSE 607
 DB 555 KTASKAHVKNKDLGNMEENKLE 575

 RESULT 6
 ID C166_CARAU STANDARD; PRT; 555 AA.
 AC Q90304;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE CD166 antigen homolog precursor (Neuroilin) (DM-GRASP homolog).
 OS Carassius auratus (Goldfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 OC Cypriniformes; Cyprinidae; Carassius.
 NC NCB1_TaxId=7957;

RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Retina;
 RA Laessing U.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RP [2]
 RP SEQUENCE OF 30-552 FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Retina;
 RX MEDLINE=94299040; PubMed=8026643;
 RA Laessing U., Giordano S., Stecher B., Lottspeich F., Stuenkel C.A.O.;
 RT "Molecular characterization of fish neurotulin: a growth-associated
 cell surface protein and member of the immunoglobulin superfamily in
 the fish retinectal system with similarities to chick protein
 DM-GRASP/SC-1/BEN";
 RT Differentiation 56:21-29(1994).
 RL
 CC -1- FUNCTION: CELL ADHESION MOLECULE. INVOLVED IN NEURITE EXTENSION BY
 CC NEURONS VIA HETEROPHILIC AND HOMOPHILIC INTERACTIONS.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: PRESENT ON ALL RETINAL GANGLION CELLS (RGCs)
 CC AND THEIR AXONS (IN EMBRYO). ABSENT FROM MATURE AXONS ALONG MOST
 CC OF THEIR LENGTH, BUT IS PRESENT ON NEW AND GROWING AXONS DERIVED
 CC FROM THE RGCs AT THE RETINAL MARGIN. REMAINS ON ADULT RGCs ONLY AT
 CC CELL-CELL CONTACT SITES AND IS CONTINUOUSLY FOUND IN THE RETINAL
 CC AXON TERMINAL ARBOR LAYERS OF THE ADULT TECTUM.
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
 CC -1- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE V-TYPE DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: L25056; AAC38015.2; -
 DR HSSP: Q13740; 1KJC.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003600; Ig_Like.
 DR InterPro: IPR003596; Ig_v.
 DR Pfam: Pf00047; Ig; 3.
 DR SMART: SM00410; Ig_Like; 3.
 DR SMART: SM00406; Ig; 1.
 DR PROSITE: PS00290; IG_MHC; 1.
 KW Signal: Developmental protein; Cell adhesion; Immunoglobulin domain;
 KW Repeat: Glycoprotein; Transmembrane.
 FT SIGNAL 1 22
 FT CHAIN 23 555
 FT DOMAIN 23 499 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 500 520 POTENTIAL.
 FT DOMAIN 521 555 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 31 117 IG-LIKE V-TYPE DOMAIN 1.
 FT DOMAIN 147 224 IG-LIKE V-TYPE DOMAIN 2.
 FT DOMAIN 256 313 IG-LIKE C2-TYPE DOMAIN 1.
 FT DOMAIN 340 392 IG-LIKE C2-TYPE DOMAIN 2.
 FT DOMAIN 419 477 IG-LIKE C2-TYPE DOMAIN 3.
 FT DISULFID 38 110 POTENTIAL.
 FT DISULFID 154 217 POTENTIAL.
 FT DISULFID 263 306 POTENTIAL.
 FT DISULFID 426 470 POTENTIAL.
 FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 171 171 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 350 350 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 441 441 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 465 465 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 555 AA; 60371 MW; 5A4A8014F0DBFF68 CRC64;

Query Match 13.8%; Score 465.5; DB 1; Length 555;
 Best Local Similarity 26.5%; Pred. No. 1,8e-23;
 Matches 167; Conservative 111; Mismatches 234; Indels 119; Gaps 27;

QY 6 LVCAFLAAC-----CCCPVAGVPGEAAPAPAEVEVEVSTALLKC--GLSGSGNLISH 59
 DB 7 LIGAFIAAAVFRPGSGVGVIGLGE-----TIYPCNDGTRKPKDG-LIF 50
 QY 60 VMFVSVHKKKRLILRRVGGOG-----SEPGYEORLSLDGRATLALTYTPDERFL 115
 DB 51 TWKVKYKDDGSPGDLVVKQAQRDEATVSATDGYKRSVISA-ANSLSLARSLDQRYFT 109
 QY 116 CQ-GKRPRSGEYRIOLRYVKAPEEPNIOVNPGLGVNSKEPE-----EVAICVGRNGVPI 169
 DB 110 CMVYSTNLEETSVKVKHKSAPYIK-----NNKLENGKLTJGECYVENANPP 162
 QY 170 PVIWYKNGRPILKEEKNRHHIOSQTVES-SGLYTLQSLTKAQLVEKDKDAQFYCELMYR 228
 DB 163 ADLWKKNNQTLVDGKTLITITSTIKKIKITLSTSSSLQYTKAREDEVSQFTC----- 217
 QY 229 LBSGNM-----KESREVTVPVYPIEKWLEVEPYGMKLEGDVREICLAGNPP---H 281
 DB 218 --TAKHVMGPDVVSSEPSPIHYPIEKVSLQVVSOSPINREGDVTLKQADGPNPPISFN 275
 QY 282 FSTSKGNPSTREAEERTTNDNGVLYLEPARKHSGRVECOGLDPTMISLSEPOELLVN 341
 DB 276 FNT-----KGGKVTYDKDYITLTGYTRADSGIYKCSLINDVMES-----TGFVYS 323
 QY 342 YVSDVVSPPADE-ROEGSSLTLTCEAESSODLEFQWLTREETGVLEKGPVLIHLDKRE 400
 DB 324 FL-DVSLFTPLGKLVKNGENLIVSLDKNASSAKATWMD--NRKLDKLP--DFSKLTYS 378
 QY 401 AGGGRYCAVASVSIPLANTOLVNAIIRPPMMA--FKRKWVAKENNVNLSCASGHP 458
 DB 379 DADLVYCDV--SIEGKISLSELTVEGIPKITSILFKRSSDGHKKV--LTCAEESP 432
 QY 459 RPIISNVNAGTSEQODQDQPVLTINLVTPPELLLETGVCYASNDLKNSTILFLELVN 518
 DB 433 KPDQVQSVNGTNDVSYNNKGATYKLTVPKSNLT--VSCLVTKKIGEDFKEI----- 483
 QY 519 LITLTPDSNTTGLSTSTASPHTRANSISTERKLPESRGGVIVAVYCIILVALVCAV 578
 DB 484 -----SVFSQKNEDGTE-----QAKVIVGIYGLVLAALVGL 516
 QY 579 LYFLY-KKGLKPCRSKQEI-TLPPSRKE 607
 DB 517 IYWIYIKTRQSGWKTEGEAGTSESKLE 547
 RESULT 7
 ID C166_BRARE STANDARD; PRT; 564 AA.
 AC Q90460; Q90480;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE CD166 antigen homolog precursor (Neurotin) (DM-GRASP homolog).
 GN CD166.
 OS Brachydanio rerio (Zebrafish) (Zebra danio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
 OC Cypriniformes; Cyprinidae; Danio.
 NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP MEDLINE=94299040; PubMed=8026643;
 RA Rankl J.P., Chang S., Kuwada J.Y.;
 RT "The molecular cloning and characterization of potential chick
 DM-GRASP homologs in zebrafish and mouse.";
 RL J. Neurobiol. 25:831-845(1994).
 RN [2]
 RP SEQUENCE OF 398-561 FROM N.A.
 RP MEDLINE=94299040; PubMed=8026643;
 RA Laessing U., Giordano S., Stecher B., Lottspeich F., Stuenkel C.A.O.;
 RT "Molecular characterization of fish neurotulin: a growth-associated
 cell surface protein and member of the immunoglobulin superfamily in
 the fish retinectal system with similarities to chick protein

RT DM-GRASP/SC-1/BEN. ;
 RL Differentiation 56:21-29(1994).
 CC -1- FUNCTION: CELL ADHESION MOLECULE. INVOLVED IN NEURITE EXTENSION BY
 CC NEURONS VIA HETEROFILIC AND HOMOPHILIC INTERACTIONS.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
 CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE V-TYPE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L25273; AAA50024.1; -
 DR EMBL: L25057; AAA50048.1; -
 DR HSSP: Q13740; IKUC.
 DR ZFIN: ZDB-GENE-990415-30; cd166.
 DR InterPro: IPR003006; IG_MHC.
 DR InterPro: IPR003600; IG_Like.
 DR InterPro: IPR003596; IG_V.
 DR Pfam: PF00047; Ig_5.
 DR SMART: SM00410; IG_Like; 3.
 DR SMART: SM00406; IGV; 1.
 DR PROSITE: PS00290; IG_MHC; 1.
 KM Cell adhesion: Immunoglobulin domain; Glycoprotein; Transmembrane;
 KM Repeat: Signal.
 FT SIGNAL 1 24
 FT CHAIN 1 25
 FT DOMAIN 25 564
 FT TRANSMEM 25 507
 FT DOMAIN 508 528
 FT DOMAIN 529 564
 FT DOMAIN 31 117
 FT DOMAIN 147 222
 FT DOMAIN 256 314
 FT DOMAIN 340 389
 FT DOMAIN 419 477
 FT DISULFID 38 110
 FT DISULFID 154 217
 FT DISULFID 263 306
 FT DISULFID 426 470
 FT CARBOHYD 92 92
 FT CARBOHYD 350 350
 FT CARBOHYD 441 441
 FT CARBOHYD 465 465
 FT CONFLICT 488 497
 FT SEQUENCE 564 AA: 61273 MW: ADAAVC293A607929 CRC64;
 Query Match 13.0%; Score 438; DB 1; Length 564;
 Best Local Similarity 26.0%; Pred. No. 1.2e-21;
 Matches 161; Conservative 112; Mismatches 254; Indels 92; Gaps 26;
 Oy 7 VCAFLIAACCCPRVAGVGEAEAPAPLVEVEGSTALLKGLSOSGNSHVDWFSVH 66
 Db 12 IAAALFAGSCLPTVIGLYGER-----IEVPCNN-----GNKKPDG-LIFTKWKYAK 57
 Oy 67 KEKRLIFRVNGGQ-----SEGEYEDRLSLQDRGATLALTYVTPQDERFLFCQ-GKRP 121
 Db 58 DDGSPGDLILKQADDPDTPVAMGQYKTRVISA-ANSLILTAOSGLTDGDRFTCMVVSST 116
 Oy 122 RSOEYRIOLRYKAPKEEPIQVNPGLGIVNSKEPEEVATCGRNGYPLPOVITWKNRPL 181
 Db 117 NLEESVEVKKHKKRPAVIR-KMYKELENGK-LTQLEGCVESNPADILIMKNNQAL 174
 Oy 182 KEKNRNVHQSQTVES-SGLTYQSLILKAQLVKEKDAQFYCEINLRPLSGNHMKESRE 240
 Db 175 VDDGKTLITSDVTKDPVTGLSTSSSRRLQYTAKEKDVASQFTC-----YAKHVTGNQ 227
 Oy 241 VTVP-----VFPTPEKVMLEVEPVGMLKGDVEYELRCLADGNPPP---HFSISKONPSTRE 293

Db 228 VSTPDTFQIRYPTKEKVSLOQVVSQSPIREDDVTLKQADGNPPTSFNFNI-----K 279
 Oy 294 ABEETDNGVLVLEPARKHESRGVEQGLDITMISLSEPPQLVINYSDVRVSPAP 353
 Db 280 GKRVYVDKDYTLTGLTVYRADSGVYKCSLDDNVMES-----TQIVVSFL-DASLTPPTGK 334
 Oy 354 E-ROEGSSLTJTCBAESSODLEFQWLEETGOVERPVQLDLKKEAGGRCVAVSP 412
 Db 335 VLKRLGELVLYSLEKNNASSEKVTWTD--NRKLDKLP--DFSQLRSDAGLYCVD--- 387
 Oy 413 SIPLNRTQLVNVAIFSPPMWA--FKERKYVKNENVLNLSCEASGHPRTISMVNGTA 470
 Db 388 SIEIKHSFSELPVLEEGPRITGLTKHRSNDGKHV---LTCEAGSGSPKEVQMSVNGTD 444
 Oy 471 SEQQDDQRVLTNLVLTVELLETGYECYASNDLGKNTSITLLELVNLTTLTPDSWTT 530
 Db 445 DETSYVNGKATYKLTLPVPSKLT--VSCLVTKRLGFDT-----KDISVPSLEEKPKRP 496
 Oy 531 GLSTSTASPHITRANSTSTERKLPESRGVYIAVYICILVAVLAVGLYFLY-KKGLKP 589
 Db 497 G-----KNEDGADQAVIYGVVGLFLAALVGLIYLIYIKKTRQG 537
 Oy 590 CRSGKQET-TLPPSRKE 607
 Db 538 SWTKETKETGTSSESKILE 556
 RESULT 8
 ICRR_DROME STANDARD; PRT; 764 AA.
 ID ICRR_DROME
 AC 008180;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Irregular chiasm C-roughnest protein precursor (IRRC protein).
 GN RST.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94102535; PubMed=7503814;
 RA Ramos R.G., Igloi G.L., Lichte B., Baumann U., Maier D.,
 RA Schneider T., Brandstaetter J.H., Froehlich A., Fischbach K.-F.,
 RT "The irregular chiasm C-roughnest locus of Drosophila, which affects
 RT axonal projections and programmed cell death, encodes a novel
 RT immunoglobulin-like protein.";
 RT Genes Dev. 7:2533-2547(1993).
 RL Genes Dev. 7:2533-2547(1993).
 CC -1- FUNCTION: REQUIRED FOR CORRECT AXONAL PATHWAY FORMATION IN
 CC THE OPTIC LOBE AND FOR PROGRAMMED CELL DEATH IN THE DEVELOPING
 CC RETINA.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: POSTEMBRYONIC EXPRESSION IS STRONG IN THE
 CC DEVELOPING OPTIC LOBE AND IN THE EYE IMAGINAL DISC.
 CC -1- DEVELOPMENTAL STAGE: STRONGLY EXPRESSED IN EMBRYOS. ALSO FOUND
 CC IN LATE LARVAL AND PUPAL STAGES.
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
 CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z21641; CAAT9756.1; -
 DR EMBL: L11040; AAA16532.1; -
 DR PIR: A49448; A49448.

DR	PIR: S34129; S34129.	
DR	HSPB: P56276; 1TK.	
DR	PLAYBASE: FBgn0003385; rsc.	
DR	InterPro: IPR003006; IG_MHC.	
DR	InterPro: IPR003598; IG_C2.	
DR	InterPro: IPR003600; IG_1like.	
DR	Pfam: PF00047; 1q; 5.	
DR	SMART: SM00410; IG_1like; 4.	
DR	SMART: SM00408; IGC2; 1.	
KW	Transmembrane; Immunoglobulin domain; Glycoprotein; Signal; Repeat.	
KW	Cell adhesion.	
FT	SIGNAL	1 19
FT	CHAIN	20 764
FT	DOMAIN	20 533
FT	TRANSMEM	534 556
FT	DOMAIN	557 764
FT	DOMAIN	32 115
FT	DOMAIN	137 221
FT	DOMAIN	245 261
FT	DOMAIN	265 332
FT	DOMAIN	353 414
FT	DOMAIN	432 515
FT	DOMAIN	637 660
FT	CARBOHYD	211 211
FT	CARBOHYD	313 313
FT	CARBOHYD	393 393
FT	CARBOHYD	400 400
FT	CARBOHYD	507 507
SQ	SEQUENCE	764 AA; 82947 MW; 262225D5D2A1C181 CRC64;

[illegible]

Db	532	--LMTVCG-----ISVVAPLVLTIL-VVVY----	555
Qy	565	KGLPCRRSGEOETLPSPRSKSELVEYKSDKLPEMGLLGSSGDKRAPDQSEKIIDL	644
Dd	556	--IKCKRRTK---LPP-----ADVISEHOITKNKGVSCKLEPDGRITSNDL	597
Qy	645	R 645	
Dd	598	K 598	
<hr/>			
RESULT 9			
PK7_CHICK	STANDARD:	PRT: 1051 AA.	
ID1_PTK7_CHICK	Q91048:		
AC	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	30-MAY-2000 (Rel. 39, Last annotation update)		
DE	Tyrosine-protein kinase-like 7 precursor (kinase like protein).		
GN	PTK OR KIG.		
OS	Gallus gallus (Chicken).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
CC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;		
OX	NCB1_TaxID=9031;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Embryonic brain;		
RX	MEDLINE=91271500; PUBMED=1711213;		
RA	Chou Y.-H., Hayman M.J.;		
RT	"Characterization of a member of the immunoglobulin gene superfamily		
RT	that possibly represents an additional class of growth factor		
RT	receptor.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 88:4897-4901(1991).		
CC	-!- FUNCTION: MAY FUNCTION AS A CELL ADHESION MOLECULE. LACKS PROBABLY		
CC	THE CATALYTIC ACTIVITY OF TYROSINE KINASE.		
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein.		
CC	-!- TISSUE SPECIFICITY: EXPRESSED IN BONE MARROW, SPLEEN, BURSA,		
CC	THYMUS AND BRAIN. WEAKLY EXPRESSED IN FIBROBLASTS. ALSO EXPRESSED		
CC	IN EMBRYONIC LIVER.		
CC	-!- SIMILARITY: CONTAINS 7 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.		
CC	-!- SIMILARITY: BELONGS TO THE INSULIN RECEPTOR FAMILY OF TYROSINE-		
CC	PROTEIN KINASAS. TRK-TYPE SUBFAMILY.		
CC	-----		
CC	This SWISS-Prot entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation		
CC	at the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@sib-sib.ch).		
CC	-----		
DR	EMBL: M63437; AAA48933.1; -		
DR	HSSP: P00523; 2PTR.		
DR	InterPro: IPR000719; Euk_pKinase.		
DR	InterPro: IPR003006; Ig_MHC.		
DR	InterPro: IPR003598; Ig_C2.		
DR	InterPro: IPR003600; Ig_Like.		
DR	InterPro: IPR001245; Tyr_kinase.		
DR	Pfam: PF00047; Ig_7.		
DR	Pfam: PF00069; pkinase; 1.		
DR	PRINTS: PR00109; TYRKINASE.		
DR	SMART: SM00410; IG_Like; 2.		
DR	SMART: SM00408; IGC2; 4.		
DR	SMART: SM00219; TYRK; 1.		
DR	PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.		
DR	Receptor: PSS0011; PROTEIN_KINASE_DOM; 1.		
KW	Immunoglobulin domain; Repeat.		
FT	SIGNAL	1 22	POTENTIAL.
FT	CHAIN	23 1051	TYROSINE-PROTEIN KINASE-LIKE 7.
FT	DOMAIN	23 685	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	686 706	POTENTIAL.

```

FT DOMAIN 707 1051 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 33 95 IG-LIKE C2-TYPE DOMAIN 1.
FT DOMAIN 130 194 IG-LIKE C2-TYPE DOMAIN 2.
FT DOMAIN 227 289 IG-LIKE C2-TYPE DOMAIN 3.
FT DOMAIN 319 379 IG-LIKE C2-TYPE DOMAIN 4.
FT DOMAIN 407 469 IG-LIKE C2-TYPE DOMAIN 5.
FT DOMAIN 498 558 IG-LIKE C2-TYPE DOMAIN 6.
FT DOMAIN 587 652 IG-LIKE C2-TYPE DOMAIN 7.
FT DOMAIN 777 1048 PROTEIN KINASE, INACTIVE.
FT DISULFID 40 88 BY SIMILARITY.
FT DISULFID 137 187 BY SIMILARITY.
FT DISULFID 234 282 BY SIMILARITY.
FT DISULFID 326 372 BY SIMILARITY.
FT DISULFID 414 462 BY SIMILARITY.
FT DISULFID 505 551 BY SIMILARITY.
FT DISULFID 594 645 BY SIMILARITY.
FT CARBOHYD 103 103 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 225 225 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 264 264 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1051 AA: 116366 MW: 1752442AE4CB702 CRC64;

```

```

Query Match 7.7%; Score 259.5; DB 1; Length 1051;
Best Local Similarity 20.1%; Pred. No. 1,7e-09;
Matches 149; Conservative 89; Mismatches 263; Indels 241; Gaps 31;

```

```

QY 29 EGPAPELVEVSGSALLKCLSGSOGMLSHVD-----WF-----SVHK 67
DB 119 KQPA-SAAELIOPSSSTVLRNC-----HIDGHPRTMQWDFDGAFLPDGKGTYSVSS 167
QY 68 EKRTLIFFVRGOGGSEGEY-----EORTLDDRGATLALQ-----Y 106
DB 168 KERLTILR--GAGPDDGLYYCARPARAGSVCSQNFILNITDESFPAVVPPEDLY 224
QY 107 TPQDRIQLCQ--GKRPSQSY-----RIQLR--VYK-- 134
DB 225 TKNEBAMPDQFAAVPRPTQEMLEFEDSPITNRKTYFANGSLILIQVKARSTGVKCIQ 284
QY 135 -----APEEPNIOVPLGIYVNSKEPEEVAATCGRNGYPIPOV 172
DB 285 HGQKGKALVYKATLRLAEIEMAFSPKSVLTANDGHV-----SCACPRGVPPTPOV 335
QY 173 IWKYK-----GRPLKEKNRHHIOSSQTVESSGLYTLOSILKAQLYEKDQAQFCYL 225
DB 336 WMRNQERVPYAGRVYQEAQLVF--TSITEADAGIYTCNAANKA----- 378
QY 226 NYRLPSGNHMKESRETVYVFYPRPEKWKLEVEFYVGMKEDRVEITGLADGNPPHFSTIS 285
DB 379 -----GEKKQELSTIV-----ATVPKWVMPKDSQLEESKPGYLCILTKASIKPTVTWY 427
QY 286 KQNSTREBAEETTNDNGVLVLEPARKHSGRECOGLDIDTMISSLSEQQLLVNY-- 342
DB 428 RGVYSIEDSRFEISENGTLRINNVEYIDGTMYK-----VSSTPASSISBYTANV 477
QY 343 --VSDVAVSAAAPERQ--EGSSLTILCEAESSQDLEFQWLKEE--TGQVLERGPVLQ 393
DB 478 HVLEKLEKFTPPQPLQCEMEFNKEVTVSCATGREKPTIOMTKDGSLSHVSHRAGILS 537
QY 394 LHDLRKRAAGGRCVAVSPISGLNRTOLYVNAIEGPPMARK--ERKWKVENNVNLINIS 451
DB 538 FHKVSRSDSNYTCIASNSPOGIRATVOIVAAVY--VTRKLEPEPTTVYQGTAFMQ 593
QY 452 CEASGAPRPTISWNVNGTASEDQDQRYLSTLNLVLPPELLETGECTASNDLKNTSI 511
DB 594 CQAEQDPVPHIQWK-----GMDKILDPSSKLLPRQIMPNSSLV--IYDTTDSKTYICI 646
QY 512 -----LFLELVNLTLLTPDSNTTGTGLSTASHTPTANSTERRKILPEDESAGV 560
DB 647 AGNSCNIKHREALFLVYVVDKPAAEDEG-----PSSHPTPKMIQT-----IGL 688

```

```

QY 561 VIVAIVYICLVAVIGAVLYFLYKKGKLPGRSGKQETILPPRSKSEL----- 608
DB 689 SVGAAYAYITII--VLG--LMFYCKR-----RRKARLKKHPEGEPEPEECNGSTGLQNG 739
QY 609 --VEYKSKLPEBEGGLQSS 628
DB 740 QTTAEIQEVALFTNLGSSSGAS 761

RESULT 10
NCAL_XENLA STANDARD: PRT: 1088 AA.
AC P16170:
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Neural cell adhesion molecule 1, large isoform precursor (N-CAM 180)
DE [contains: N-CAM 140].
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=83355;
RN [1]
RP SEQUENCE FROM N. A.
RX MEDLINE=90098771; PubMed=2481269;
RA Kriegl P.A., Sakaguchi D.S., Kintner C.R.;
RT "Primary structure and developmental expression of a large
RT cytoplasmic domain form of Xenopus laevis neural cell adhesion
RT molecule (NCAM).";
RL Nucleic Acids Res. 17:10321-10335(1989).
CC -1- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
CC NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
CC NEURITES, ETC.
CC -1- SUBCELLULAR LOCATION: TYPE I membrane protein.
CC -1- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF
CC N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN NEURON AND IN PRESUMPTIVE NEURAL
CC TISSUE.
CC -1- DEVELOPMENTAL STAGE: THE MRNA ENCODING THIS LD-NCAM IS THE MAJOR
CC TRANSCRIPT PRESENT IN BOTH MATERNAL RNA AND IN THE EMBryo DURING
CC EARLY NEURAL DEVELOPMENT.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DB EMBL: M25696; AAA49909.1; -
DR PIR: S09600; IXYLNL.
DR HSSP: P56276; IYLR.
DR InterPro: IPR003361; FN_III.
DR InterPro: IPR003006; IG_MHC.
DR InterPro: IPR003598; IG_C2.
DR Pfam: PF00041; fn3; 2.
DR Pfam: PF00047; Ig; 5.
DR SMART: SM00060; FN3; 2.
DR SMART: SM00408; IGc2; 5.
KW Cell adhesion; glycoprotein; Transmembrane; Repeat; Brain;
KW Immunoglobulin domain; Alternative splicing; Signal.
FT CHAIN 1 19
FT SIGNAL 19
FT DOMAIN 20 1088 NEURAL CELL ADHESION MOLECULE 1, LARGE
FT ISOPFORM.
FT TRANSMEM 706 723 EXTRACELLULAR (POTENTIAL).

```

```

FT DOMAIN 724 1088 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 34 100 IG-LIKE C2-TYPE DOMAIN 1.
FT DOMAIN 129 193 IG-LIKE C2-TYPE DOMAIN 2.
FT DOMAIN 225 289 IG-LIKE C2-TYPE DOMAIN 3.
FT DOMAIN 316 386 IG-LIKE C2-TYPE DOMAIN 4.
FT DOMAIN 413 480 IG-LIKE C2-TYPE DOMAIN 5.
FT DOMAIN 512 589 FIBRONECTIN TYPE-III 1.
FT DOMAIN 618 686 FIBRONECTIN TYPE-III 2.
FT DOMAIN 149 153 HEPARIN-BINDING (POTENTIAL).
FT DOMAIN 158 162 HEPARIN-BINDING (POTENTIAL).
FT DISULFID 41 93 PROBABLE.
FT DISULFID 136 186 PROBABLE.
FT DISULFID 232 282 PROBABLE.
FT DISULFID 323 379 PROBABLE.
FT DISULFID 420 473 PROBABLE.
FT CARBOHYD 82 219 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 219 219 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 310 310 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 341 341 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 443 443 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 472 472 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPIC 804 1049 MISSING (IN ISOBORN N-CAM 140).
SQ SEQUENCE 1088 AA; 117778 MW; 6273855803F3E83 CRC64;

```

Query Match 7.6%; Score 256; DB 1; Length 1088;
 Best Local Similarity 18.1%; Pred. No. 3,1e-09;
 Matches 155; Conservative 100; Mismatches 266; Indels 336; Gaps 35;

```

QY 33 PELVEVEGSTALLKCGLSQGNLSHVDFVYHKEKRLIFRVROGQSGEYEQRL 92
DB PDGEISLGEKFLC---QVSGEATDISWSPTEKLV-----QOQI 66
QY 93 SL---QORGTALQVTPQDERIFLC-----QGR-----RPRQRY 126
DB SVRSDDYTLTLTYNASSODAGIKVCASNAEGSEGTYNLKITOKLFFKNAPPPQEF 136
QY 127 R-----IOLRYKAPPE-----138
DB KEGEDAVIICDVSSISITWRHKDVFKKDVRVYLANNTYLOIRGIKTKDEGYRC 186
QY 139 -----PNIQVPLGIPVNSKEPEVATCVGRNGYPIPOVITWK 176
DB EGRILARGEINVKDIQYIVNVPPIQARQVRVATNAMESVYLCADGFPDEISMLK 246
QY 177 NGRPLKEKNVHIQSSQYVSSSLYLOSLKRAQLYKEDKDAQFYELMYRLPSGHHM 236
DB KGEPIEDGEEEKISFNE-----QSEMTIHHKEKDEMEVSCIAN-----NOAG 289
QY 237 ESREVTVPVPEYREKVMLEVEPVGMLEKGRVETRCIADGNPPPHFISIKQNPSTRAEE 296
DB EAETITLLKYAKKITYVENKTAIV--ELDEITITCEASDPPIP--SITWRTAVRNISSE 345
QY 297 ETTDNGVLYLEPARKK-----HSGRYEC-----QGLDIDIMISLSEPOELVN 341
DB ATTLDGHIYVKEHTRMSALTLKDIQYTDAGEYFCIASNPICGVDMQAMV-----FEVQ 397
QY 342 YVSDVRSAPAPERQEGSSLITLCEASSODLEFOWLREFEGVLE-----RGPV 391
DB YAPKIR-GPVVYVYTWEGNPNVITCEVAFHAPRAVTPFRD--GQLPSSNSNIKIYSGPT 434
QY 392 ---LQHLDLKREAGGYRCV-----ASVPSIPGLANTOLVN--VAIFG 429
DB SSSLEVNDSDNDGNVNTAINTIGHSEFELIYQADTPSSPAIRKVEVYSSVVMIVFD 514
QY 430 PP-----WMAEKERVWK-----ENM--VINLSCEAS-----435
DB EPDSTGVPIIKYKAERKRVIGHEKMHKYYDAKEVNAESIITYWGLPETSVMYKLSAMN 574
QY 456 ---GHPRTISMVNV-----GTASE-----ODOPQPVV-----LST 483
DB 575 GKGLDSTPSSQEFITTPYVREPSAKLVGHLSEDSNSTIKVILKODDGGSPIRHYLVYRA 634

```

```

QY 484 LNLVY-TPEL-----LETGVE---CYASNDLCKNTSIL--FLELVNLTTLT 523
DB 635 LNALEKPEKMRPSSSHHMYLKALEMNDYEVYVAENQGRKALLSPFTAKPATY 694
QY 524 PSDNTTGLSTASTASPHRANSTRNTERKLPEPSRGVYIVAVICILAVLAVLYFLY 583
DB 695 ATASAGTGLGTG-----AIVGILIVIPVLLLVVDVYVYCEFLN 731
QY 584 KKGKLPK-----RSGKQETLTPPSKSELVEYVSKDKLPEEGGLLOG 626
DB 732 KCGLLMCIANVNCGRAGPAKAGKDIEEGKAAPSKDESK--EPIYEVRE-----778
QY 627 SSGDKRAPDQGEKYID 643
DB 779 ---EERTPHHDSNQIE 792
RESULT 11
NCB2_MOUSE
ID P13594: 061950: STANDARD; PRT; 725 AA.
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Neutral cell adhesion molecule, phosphatidylinositol-linked isoform
DE Precursor (N-CAM 120) (NCAM-120).
GN NCAM1 OR NCAM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=87246524; PubMed=3595563;
RA Barthele D., Santoni M.-J., Wille W., Ruppert C., Calix J.-C.,
RT Hirsch M.-R., Fontecilla-Camps J.-C., Goridis C.;
RT "Isolation and nucleotide sequence of mouse NCAM cDNA that codes for
RT a Mr 79,000 polypeptide without a membrane-spanning region.";
RL EMBO J. 6:907-914(1987).
RN [2]
RP SEQUENCE OF 20-700 FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=89251563; PubMed=2721486;
RA Santoni M.J., Barthele D., Vopper G., Boned A., Goridis C., Wille M.;
RT "Differential exon usage involving an unusual splicing mechanism
RT generates at least eight types of NCAM cDNA in mouse brain.";
RL EMBO J. 8:385-392(1989).
RN [3]
RP SEQUENCE OF 642-725 FROM N.A.
RX MEDLINE=88283628; PubMed=3396534;
RA Barbas J.A., Chaix J.C., Steinmetz M., Goridis C.;
RT "Differential splicing and alternative polyadenylation generates
RT distinct NCAM transcripts and proteins in the mouse.";
RL EMBO J. 7:625-632(1988).
RN [4]
RP SEQUENCE OF 20-36.
RX MEDLINE=86140120; PubMed=3512556;
RA Rougon G., Marshak D.R.;
RT Structural and immunological characterization of the amino-terminal
RT domain of mammalian neural cell adhesion molecules.";
RL J. Biol. Chem. 261:3396-3401(1986).
CC -!- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
CC NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
CC NEURITES, ETC.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI anchor.
CC -!- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF
CC N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -!- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----

```


CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: Y00051; CAA68263.1; -.
DR EMBL: X15049; CAA33148.1; ALT_SEQ.
DR EMBL: X07195; CAA30173.1; -.
DR PIR: A29673; IJMSNG.
DR MGD: MGI:97281; Ncam.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003598; Ig_C2.
DR Pfam: PF00041; fn3; 2.
DR Pfam: PF00047; Ig; 5.
DR SMART: SM00060; FN3; 2.
DR SMART: SM00408; IgC2; 5.
KW Cell adhesion; Glycoprotein; Repeat; Alternative splicing;
KW Immunoglobulin domain; Signal; Heparin-binding; GPI-anchor.
FT SIGNAL 1 19
FT CHAIN 20 725
FT DOMAIN 34 103 NEURAL CELL ADHESION MOLECULE,
FT DOMAIN 132 196 PROSPATIDYLINOSITOL-LINKED ISOFORM.
FT DOMAIN 228 295 IG-LIKE C2-TYPE DOMAIN 1.
FT DOMAIN 323 393 IG-LIKE C2-TYPE DOMAIN 2.
FT DOMAIN 420 487 IG-LIKE C2-TYPE DOMAIN 3.
FT DOMAIN 519 596 IG-LIKE C2-TYPE DOMAIN 4.
FT DOMAIN 625 692 IG-LIKE C2-TYPE DOMAIN 5.
FT DOMAIN 152 156 FIBRONECTIN TYPE-III 1.
FT DOMAIN 161 165 FIBRONECTIN TYPE-III 2.
FT DISULFID 41 96 HEPARIN-BINDING (POTENTIAL).
FT DISULFID 139 189 HEPARIN-BINDING (POTENTIAL).
FT DISULFID 235 288 PROBABLE.
FT DISULFID 330 386 PROBABLE.
FT DISULFID 427 480 PROBABLE.
FT CARBOHYD 222 222 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 348 348 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 450 450 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 479 479 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 261 268 ERSRVS -> DEKIFSD (IN REF. 2).
FT CONFLICT 273 273 V -> L (IN REF. 2).
FT CONFLICT 354 355 QD -> KT (IN REF. 2).
FT CONFLICT 349 349 T -> K (IN REF. 2).
FT CONFLICT 572 572 T -> R (IN REF. 2).
FT CONFLICT 575 575 D -> V (IN REF. 2).
FT CONFLICT 589 594 MOPSSES -> SAATEF (IN REF. 2).
FT CONFLICT 600 602 PEL -> REP (IN REF. 2).
FT CONFLICT 657 657 H -> D (IN REF. 2 AND 3).
SQ SEQUENCE 725 AA: 80296 MW: C2AE8B84461C6B2F CRC64;

Query Match 7.3%; Score 246.5; DB 1; Length 725;
Best Local Similarity 20.6%; Pred No. 7.4e-09;
Matches 141; Conservative 92; Mismatches 261; Indels 189; Gaps 33;

QY 33 PELVEVGSSTALNCGLSQSGNLSHVD--WFSVHKERKTLIFRVGQSGSEGEYEQ 90
DB 26 POGSEISVESKEFLC---QVAGDAKKDIDSWFSPNGEKLS-----PNOQ 67
QY 91 RLSL---QDRGATLALTQVTPQDERIFLC--QGRPRRSQERYIQLRYKAPREPIQNP 145
DB 68 RLSTVWNNDDSTLTITYNANIDAGIKCVYAEQGTQSEAVNKKIRQ-----KLMEKN 122
QY 146 LGIPVNSKEPEE-VATGVCGRNGYPIPOVIWYKNGRLKEEKN-RVHISSQTSQVSSGLYT 203
DB 123 APTPOEFKGEDEAVIYCDVYVSLP-PTIIMKHKGDDVILKKDVRITVLSNNYLQIRG--- 178

QY 204 LOSILKAQVKEKRDADQFCYCELYNRLPSSGNHMKESREVTVPFYETKEVLEVEPYGMLK 263
DB 179 : : : : : IKTBEGYRCE-----GRILANG-----INKKDQIVVNNPPYQAR 217
QY 264 E-----GDRVEIRCLADGNPPPHFSISKQ-NPSTRAEER---TTNDGCVLLEP 309
DB 218 : : : : : QSIYATANLQOSTVLVCDADGFEPEPTMSWTWKDEPIENEEDERSRSVSDESSEVIRN 277
QY 310 ARKESGRGECQG-----LDIDTMISLSPEOBLLVNVSDRVSPAAREGGSSTLT 363
DB 278 VDKNDEATYVCIANKGGEODASIHLYFAFK---ITVE-----NOTAMELE--QVTL 328
QY 364 TCSEASSQDLFEQWLREET-----GOYLERGPV-----LOLHDLKREAGGRCV 408
DB 329 TCSEASGDPISITW-RITSTRNISSEEDDGHVYVSHARVSSLTLSIQYRAGEYMC 387
QY 409 AS-----VSIPLGLNRTQLVNVAIFPPPMARERKRVYKEMVNLSC 452
DB 388 ASNTIGDSQSIDLEFOYAPKIQG-----PVAVY-----TW--EGNOVAITC 427
QY 453 EASGHPRTISWNVNGTASBEDDQDQRYVLSLTVLP--ELLEET-----GVECTA 501
DB 428 EVFATPSATISWFRDQGL-----PSSNYSNIKIYNTPSASYLEVYDPSNDPQNTCTA 482
QY 502 SNDGKNTSILFELVNLTLTPDSNTTGTGSTASPHTRANSTERTKLPESRGV 561
DB 483 VNRIGQES-----LEFIVQADTPSSP-----SIDRVEPY-----SSTNQVDFDEPATGV 529
QY 562 IVANYIVCL-----VLAIVGAVLYFLYKKKLLPCRRSGKOET 599
DB 530 PILRYKAEMKSLGEESWHTFYDAKEANMGIVIMGLKDETTYSDRLAALANGGLGEIM 589
QY 600 LPSPKSELVYEVRSDKLPREMG 622
DB 590 QPSEKTIQPVPELSAPKLEGQMG 612

RESULT 12
VCAL RAT STANDARD; PRT; 739 AA.
ID VCAL RAT
AC P29534;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Vascular cell adhesion protein 1 precursor (V-CAM 1).
GN VCAM1 OR VCAM-1.
OS Rattus norvegicus (Rat).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CC NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=92181437; PubMed=1371918;
RA Hession C., Moy P., Tizard R., Chisholm P., Williams C., Wisk M.,
RA Burkly L., Miyake K., Kincade P., Lobb R.;
RT Cloning of murine and rat vascular cell adhesion molecule-1.;
RT Biochem. Biophys. Res. Commun. 183:163-169(1992).
CC -!- FUNCTION: IMPORTANT IN CELL-CELL RECOGNITION. APPEARS TO FUNCTION
CC IN LEUKOCYTE-ENDOTHELIAL CELL ADHESION. INTERACTS WITH THE BETA-1
CC INTEGRIN VLA4 ON LEUKOCYTES, AND MEDIATES BOTH ADHESION AND SIGNAL
CC TRANSDUCTION. THE VCAM1/VLA4 INTERACTION MAY PLAY A
CC PATHOPHYSIOLOGIC ROLE BOTH IN IMMUNE RESPONSES AND IN LEUKOCYTE
CC EMIGRATION TO SITES OF INFLAMMATION.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: EXPRESSED ON INFLAMED VASCULAR ENDOTHELIUM, AS
CC WELL AS ON MACROPHAGE-LIKE AND DENDRITIC CELL TYPES IN BOTH NORMAL
CC AND INFLAMED TISSUE.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -!- SIMILARITY: CONTAINS 7 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: M84488; AAA42332.1; .
PIR: J50675; J50675.

HSP: P19320; YVCA.
InterPro: IPR003987; ICAM_YCAM-1.

InterPro: IPR003006; I9_MHC.
InterPro: IPR003598; I9_C2.

InterPro: IPR003600; I9_Like.
InterPro: IPR003989; YCAM-1.

Pfam: PF00047; I9_5
PRINTS: PRO1472; ICAMVCA1.

PRINTS: PRO1474; VCAM1.
SMART: SM00410; IC_Like; 1.

SMART: SM00408; IGC2; 4.
Immunoglobulin domain; Glycoprotein; Cell adhesion; Transmembrane;

Repeat; Signal. 1. 24. PROBABLE. VASCULAR CELL ADHESION PROTEIN 1.
CHAIN 25 698 EXTRACELLULAR (POTENTIAL).

TRANSMEM 721 739 POTENTIAL. CYTOPLASMIC (POTENTIAL).

DOMAIN 38 91 IG-LIKE C2-TYPE DOMAIN 1.
DOMAIN 129 198 IG-LIKE C2-TYPE DOMAIN 2.

DOMAIN 237 287 IG-LIKE C2-TYPE DOMAIN 3.
DOMAIN 326 379 IG-LIKE C2-TYPE DOMAIN 4.

DOMAIN 418 496 IG-LIKE C2-TYPE DOMAIN 5.
DOMAIN 525 575 IG-LIKE C2-TYPE DOMAIN 6.

DOMAIN 612 675 IG-LIKE C2-TYPE DOMAIN 7.
DOMAIN 723 773 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 531 531 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 561 561 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 650 650 N-LINKED (GLCNAC. . .) (POTENTIAL).

SEQUENCE 739 AA; 81246 MW; 5C08E5A1A1B100C CRC64;

Query Match 7.3%; Score 246.5; DB 1; Length 739;
Best Local Similarity 21.5%; Pred. No. 7.6e-09;
Matches 133; Conservative 83; Mismatches 182; Indels 221; Gaps 30;

39 EWGSTALLKCGLSQSGNLSHVDFSVHKEKTLIFRVRQGGSGSEPEYEQRLSLDRG 98
38 QIGDSMLTSCSTTGCS--PSFSW-----RTQIDSPUNGKVKTEGAK----- 77
99 ATLALQVTPQDERIFLC-----OGKRPQSGEYRIOLRVYKAPPEPNIOVN-PL--GIPV 150
78 SVLTMPVSEFNEHSLTCAATNSGRLRG-----IOVDIYSPKRPDELQFSGPLEVGKPV 133
151 NSKEPEEAVWCVGRNGYPIP--QVIWYKNGRLPKKEKNVHIIQSSQYVSSGLYLOSL 208
134 MKV-----CLAPDYPIIDLEIELFKGDRMLKKQ--DEVDMARKSLTKSLVYFT-- 183
209 KAOLVEDKDAQFYCELNRIRISGNHMKESREVT--VPFYFTEKVMLEVEPVGMKEGD 266
184 ---PVIEDIEKALVCRAKLYIDQTDIPKEREVRELQVYTSKNTNEISVHSTRHEGA 240
267 RVEIRCLADGNPPPHFSIKONPSTREAEETNDNGV-----LVLEPAKHEHSR 317
241 AVMTTCASEGLPAPELFMSKK-----LDNGVLIQLSGNATLTILIAMRMDSGI 288
318 YECGIDL---DTMISLSEPEOLLVNYSDVRVPAAP--EQGESSLTITCEASSSD 372
289 YVCGVNLVGRDKTEVLELVQGEKFTVD-----ISGSOVAANOVDGVLTCAAVGCD 342
373 LEFQWLREET---GOVLENG----- 389
343 PSFSW-RTQIDSPUNGKVKTEGAKSTLTLSPVGVEDEHSLCTVTCORRIEKTIOVEY 401

390 -----PY-----IQLDHDKRE----- 400
402 SPEDPEIETISGLVGRPTVNCVPPPHLIELLKTLINKFLREIGTKSL 461
401 -----ACGGRVCA-----SVSPISGLNRTQVYVAFGPMAFKE 437
462 ETKSEMTFIPFAEDTGKALVCLAKLHSSQSEPEKOROSTQRLVYVNA---P---KE 513
438 RKVAV-----KENMYNLSCASGHPRTISWN---VNGTASEODODPQRLVSTLVIV 488
514 PTIVSPSPVPEEGSPVNLTCSSGFTPTKILMSROLKNEQLPLSQN-----TTLFMA 568
489 TPPELLTGV-ECTASNDLG 506
569 T-KMEDSGIYCEGINEAG 586
RESULT 13
NCAL_RAT NCAL_RAT STANDARD: PRT: 858 AA.
AC P13596:
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Neutral cell adhesion molecule, 140 kDa isoform precursor (N-CAM 140)
DE (NCAM-140).
OS Rattus norvegicus (rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
NP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=88059265; PubMed=3680385;
RA Small S.J., Shull G.E., Santoni M.-J., Akeson R.;
RT Identification of a cDNA clone that contains the complete coding
sequence for a 140-kD rat NCAM polypeptide";
RL J. Cell Biol. 105:2353-2345(1987).
RN [2]
RP SEQUENCE OF 355-364 FROM N.A.
RX MEDLINE=90166485; PubMed=2483093;
RA Small S.J., Haines S.L., Akeson R.A.;
RT Polypeptide variation in an N-CAM extracellular immunoglobulin-like
fold is developmentally regulated through alternative splicing";
RL Neuron 1:1007-1017(1988).
CC -1- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
NEURITES, ETC.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF
N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR SMART: SM00408; IGC2; 5.
 KW Cell adhesion; glycoprotein; Transmembrane; Repeat;
 FT Immunoglobulin domain; Alternative splicing; Signal; Heparin-binding.
 FT SIGNAL 19
 FT CHAIN 20 858 NEURAL CELL ADHESION MOLECULE, 140 KDA
 FT DOMAIN 20 721 ISOFORM.
 FT TRANSMEM 722 739 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 740 858 POTENTIAL.
 FT DOMAIN 740 858 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 132 196 IG-LIKE C2-TYPE DOMAIN 1.
 FT DOMAIN 132 196 IG-LIKE C2-TYPE DOMAIN 2.
 FT DOMAIN 228 295 IG-LIKE C2-TYPE DOMAIN 3.
 FT DOMAIN 323 403 IG-LIKE C2-TYPE DOMAIN 4.
 FT DOMAIN 430 497 IG-LIKE C2-TYPE DOMAIN 5.
 FT DOMAIN 514 615 FIBRONECTIN TYPE-III 1.
 FT DOMAIN 616 712 FIBRONECTIN TYPE-III 2.
 FT DOMAIN 152 156 HEPARIN-BINDING (POTENTIAL).
 FT DOMAIN 161 165 HEPARIN-BINDING (POTENTIAL).
 FT DISULFID 41 96 BY SIMILARITY.
 FT DISULFID 139 189 BY SIMILARITY.
 FT DISULFID 235 288 BY SIMILARITY.
 FT DISULFID 330 396 BY SIMILARITY.
 FT DISULFID 437 490 BY SIMILARITY.
 FT CARBOHYD 222 222 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 348 348 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 434 434 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 460 460 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 489 489 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 858 AA; 94658 MW; EA1A064EA050F6 CMC64;

Query Match 7.3%; Score 244.5; DB 1; Length 858;
 Best Local Similarity 20.9%; Pred. No. 1.3e-08;
 Matches 147; Conservative 92; Mismatches 245; Indels 219; Gaps 34;
 33 PELVEVEGSTALLKCGLSQSGNSLHVD--WFSVHKERKRIIFVRGOGQSGSEGEFEQ 90
 26 PSGGEISVGESEKFFLC---QVAGDAKDKDISFSPNGEKLS-----PNOQ 67
 91 RSL---QDRGATLALTVTPDDEKFLC--QGRPRSGEYRIOLRYKABEERNIQVNP 145
 68 RISVYVNDSSSTLTYYNANIDAGIKCVYAEEDGTOSEATVNVKIQ-----KLMFKN 122
 146 LGIPVNSKEPEE-VATCVGRNGYPIPOVIWYKNGRPLKEEN--NVHIOSSGVSSGLYT 203
 123 APTPOEFKGEDAVIVCVVSSLP-PTIIMKHKGHDVILKADVRFVLSNNYIQIRG--- 178
 204 LOSILKAOLVKEDKADQRYCELNRLPFGSNHMKESREVTYVFEYTERKVMLEVPVGMK 263
 179 -----IKRTDEGTYRCE-----GRILARGE-----INFKDIQIVVNPPTVQAR 217
 264 E-----GDRVEIRCLADGNPPHHSISKPNSTREAEEL-----TNDGCVLLEP 309
 218 OSIYVATNLTQSVTLVCDADGFPETPMYSWKDEPIENEDEDEKHIIFDDSESLIIRN 277
 310 ARKHEISGREGQG-----IDLDTMISLSEPOELIVNVYSDVRSVPAEREGSSSLT 363
 278 VDKNDDEAYVCIARNKAGEQDASIHKLKVPKPK---ITTYE-----NOTAMELEP--QVTL 328
 364 TCEAES-----SODLEFQWLAREELGQVLEGPVQLH-----DLKR 399
 329 TCEASGDPISITRTSTRTNISSEKASWTRPKOETLDHMYVVRSHARVSLTKSIOY 386
 400 EAGGGRCVAS-----VPSIPGLNRTQLVNVAIFGRPMMAFERKRVVWK 443
 389 TDAGEYICTASNTIGDSQSMYLEVQAPRLQG-----PVAAY-----TW-- 428
 444 ENNVLTNLSCEASGHPRTISWNVNGTASBEDODPOVRLSTLVLYTP--ELLEET----- 495
 429 EGNVLTICEVFAVPSATISWFRDQGL-----PSSNTYNIKIYVNPASVYLEVPPDSN 483
 496 ---GVECTASNDLGKMTSILFLVNLVLTLLTPDSNTTTLSTSTASPHRTANSTSTERKL 552

DB 484 DFGNVCATVANNIGDS-----LEFLTVQADIPSSP-----SIDREVEY-----SSTAQVCP 530
 OY 553 PEPEERGVYVAV-----IYCVLTV-----AVIGAVY 579
 DB 531 DEPEATGVPILTKYKAEMKSLGEEAMHSKWDAKEANMEGIIVTIMGKLPRTAVRLAAL 590
 OY 580 YFLYKRGKLPKRSGKOITLPSRKSEIYEVKSKDLPEEMG 622
 DB 591 -----NGK-----GLGEISATETFKTPVREPSAPRLQGMG 622
 RESULT 14
 PGBM_MOUSE
 ID PGBM_MOUSE STANDARD; PRT; 3707 AA.
 AC 005793;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Basement membrane-specific heparan sulfate proteoglycan core
 DE protein precursor (HSPG) (Perlecan) (PLC).
 GN HSPG2.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Melanoma;
 RX MEDLINE=92078153; PubMed=1744087;
 RA Noonan D.M., Horiig E.A., Ledbetter S.R., Vogel G., Sasaki M.,
 RA Yamada Y., Hassell J.R.,
 RT "The complete sequence of perlecan, a basement membrane heparan
 RT sulfate proteoglycan, reveals extensive similarity with laminin A
 RT chain, low density lipoprotein-receptor, and the neural cell adhesion
 RL molecule.";
 RL J. Biol. Chem. 266:22939-22947(1991).
 RN [2]
 RP SEQUENCE OF 940-1601 AND 1870-2600 FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=89034110; PubMed=2972708;
 RA Noonan D.M., Horiig E.A., Ledbetter S.R., Vogel G., Sasaki M.,
 RA Yamada Y., Hassell J.R.,
 RT "Identification of cDNA clones encoding different domains of the
 RT basement membrane heparan sulfate proteoglycan.";
 RL J. Biol. Chem. 263:16379-16387(1988).
 CC - FUNCTION: THIS PROTEIN IS AN INTEGRAL COMPONENT OF BASEMENT
 CC MEMBRANES. IT IS RESPONSIBLE FOR THE FIXED NEGATIVE ELECTROSTATIC
 CC CHARGE AND IS INVOLVED IN THE CHARGE-SELECTIVE ULTRAFILTRATION
 CC PROPERTIES. IT INTERACTS WITH OTHER BASEMENT MEMBRANE COMPONENTS
 CC SUCH AS LAMININ AND COLLAGEN TYPE IV AND SERVES AS AN ATTACHMENT
 CC SUBSTRATE FOR CELLS.
 CC - SUBUNIT: PURIFIED PERLECAN HAS A STRONG TENDENCY TO AGGREGATE IN
 CC DIMERS OR STELLATE STRUCTURES.
 CC - TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES.
 CC - PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
 CC AND O-LINKED OLIGOSACCHARIDES.
 CC - SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
 CC - SIMILARITY: CONTAINS 3 LAMININ EGF-LIKE DOMAINS.
 CC - SIMILARITY: CONTAINS 15 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC - SIMILARITY: CONTAINS 3 LAMININ G-LIKE DOMAINS.
 CC - SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC - SIMILARITY: CONTAINS 1 SEA DOMAIN.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation.
 CC The European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).


```

OY 17 CCRVAGV-----GEAEPAPELVEVEY-----GSTALLKCGLSQSGNLSHYDMF 63
DB 1653 CAGYEGEDNVGCGKCPQTESLEVOIHPSRSVPQGGPHSLRC---QVSGSPHFYFW 1709
OY 64 SVHKERKTLIFRVRGOGQSEGEYEQRLSLQDRGATLALQVTPQDERIFLCOGKR-PR 122
DB 1710 S-----REDGRPLPSSAQOR---HQSELIHFPSVQPSDAGVYICTRMLIH 1752
OY 123 SOEYRIQLRVYKAPPEPNIQVPLGIVNKSKEPEVA-----TCVGRNGYPIQVYIW 175
DB 1753 TSSRAELLYVAPSK-----PIWYVEQROQSVPBPADYFICTAKSKSPATLVMT 1806
OY 176 K--NGRPLKEEKNRHHVITSSQTVSSGLYLSQSLAQLVKEDKAQRY-CELN--YRLP 230
DB 1807 RLNGK-----LPSRAMDFNGLITIRV-----OPSDAGVYCTGSMFRAMD 1848
OY 231 SGN---HMKESREVTVPV--FYPTKEVYLEVEPVGMLKESDRVEIRCLADGNPPHF-- 282
DB 1849 QGTATLHVQVSGSTAPVASIHPQ---LIVQP-----GQQAEPKCATGNPTMLEWI 1899
OY 283 -STSKQNPSTRAEEETTNDGCVTLVPARKHSGRYEGOGLD-----LDTMISLL--S 333
DB 1900 GGFSGGLPAKQIHH-----NGIURLAIPRSDQGYLGRALSSAGQVYARMMQVHGGS 1953
OY 334 EPEELLVNVSDYRVSPPAPRQSSSLTTCABESSQLEFQWLRRETCGYLER----G 389
DB 1954 GPR-----VQVSPERTVHSGRTVRLYCRAGVPSASITW--RKEGSLPRHOAHG 2003
OY 390 PVLQHLDKREAGGGRGVASVPSIFGLNFTQLVNV-----AIFGPPMAFKERKRV 441
DB 2004 SRLRLHMSVADSGEYVCVRAN--NNIDQETSIMISVSPSTNSPPAPASPARIRIESSSR 2062
OY 442 VKENNVNLNCSGASHPPPTSMVNVG 468
DB 2063 VAEQGTLDLNCVPGHAAQYTWKRG 2089

RESULT 15
NCAL_BOVIN STANDARD; PRT; 853 AA.
ID NCAL_BOVIN STANDARD; PRT; 853 AA.
AC P31836;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Neural cell adhesion molecule, 140 kDa isoform precursor (N-CAM 140)
DE (NCAM-140).
GN NCAM OR NCAM.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Kumlantlia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Brain cortex;
RX MEDLINE=89378239; PubMed=2776887;
RA Lipkin V.M., Khramtsov N.V., Andreeva S.G., Moshnyakov M.V.,
RA Petukhova G.V., Rakitina T.V., Reschenko E.A., Ishchenko K.A.,
RA Mirzoeva S.F., Chernova M.N., Dranytsyna S.M.;
RT "calmodulin-independent bovine brain adenylate cyclase. Amino acid
RT sequence and nucleotide sequence of the corresponding cDNA.";
RU FEBS Lett. 254:69-73(1989).
RN [2]
RP SEQUENCE OF 20-36.
RX MEDLINE=86140120; PubMed=3512556;
RA Rougon G., Marshak D.R.;
RT "Structural and immunological characterization of the amino-terminal
RT domain of mammalian neural cell adhesion molecules.";
RL J. Biol. Chem. 261:3396-3401(1986).
RN [3]
RP IDENTIFICATION AS N-CAM.
RX MEDLINE=92111748; PubMed=1765159;
RA Premont R.T.;

```

```

RT "A bovine brain cDNA purported to encode calmodulin-insensitive
RT adenylate cyclase has extensive identity with neural cell adhesion
RT molecules (N-CAMs).";
RL FEBS Lett. 295:230-231(1991).
CC -1- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
CC NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
CC NEURITES, ETC.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF
CC N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -1- CAUTION: WAS ORIGINALLY (REF.1) THOUGHT TO BE A CALMODULIN-
CC INDEPENDENT ADENYLATE CYCLASE.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.1sb-sib.ch/announce/
CC or send an email to license@1sb-sib.ch).
CC -----
DR EMBL: X16451; CA34470.1; -.
DR PIR: A32976; IJBONC.
DR HSP: P40189; IBDU.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR003006; IG_MHC.
DR InterPro: IPR003598; IG_C2.
DR Pfam: PF00041; FN3; 2.
DR Pfam: PF00047; IG; 5.
DR SMART: SM00060; FN3; 2.
DR SMART: SM00408; IG2; 5.
DR K cell adhesion; glycoprotein; Transmembrane; Repeat.
KW Immunoglobulin domain; Alternative splicing; Signal.
FT STGNAL 1
FT CHAIN 20 853
FT FT 719
FT TRANSMEM 737
FT DOMAIN 738 853
FT DOMAIN 34 103
FT DOMAIN 132 196
FT DOMAIN 228 293
FT DOMAIN 321 401
FT DOMAIN 428 495
FT DOMAIN 527 604
FT DOMAIN 633 700
FT DOMAIN 152 156
FT DOMAIN 161 165
FT DOMAIN 161 165
FT DISULFID 139 189
FT DISULFID 235 286
FT DISULFID 328 394
FT DISULFID 435 488
FT CARBOHYD 222 222
FT CARBOHYD 314 314
FT CARBOHYD 346 346
FT CARBOHYD 432 432
FT CARBOHYD 458 458
FT CARBOHYD 487 487
SQ SEQUENCE 853 AA; 93893 MW; E12FD49231A368 CRC64;

Query Match 7.2%; Score 242.5; DB 1; Length 853;
Best Local Similarity 20.8%; Pred. No. 1.7e-08;
Matches 146; Conservative 93; Mismatches 245; Indels 217; Gaps 34;
OY 33 PELVEVSGSTALLKCGLSQSGVLSHYD--WESYKKEKRTLIFRVRGOGQSEGEYEQ 90
DB 26 PSQGLISGSKFFLC---QVAGAKDKDISWSPGKELI-----PNDQ 67

```

QY 91 RLST---QDRGATLALQVTPDERIFLC--QGRPRSQEYRIQLRYKAPBEPNIQVNP 145
 Db 68 RISVVMNDSSSTLTIYANIDAGIYKCVTAEDGTSEATYVNAKIFQ-----KLMFKN 122
 QY 146 LGIPVNSKEPBE-VATCVGNKGPPIPOVYWKNGRPLKEKN-RVHIQSQIYESSGLYT 203
 Db 123 APPEQEPREGEDAVYCDVSSLP-PTIIKHKGRDVIILKDVFTVLFTNNYLQIRG--- 178
 QY 204 LQSLIKQLVKEDKDAQFYCELVNRLPSGNHMKESREVTVPVPTKEKWLVEVEPYGMLK 263
 Db 179 -----IKKIDEGYRCE-----GRILARGE-----INFKDIQYIVNVPPTVQAR 217
 QY 264 E-----GDRVEINCLADGNPPHFSISKONPSTREAEET---TNDNGVLVLEPAR 311
 Db 218 QSTYNATANAGQSTVLVCNAGFPEPTVSWTKGEOJENEDKYLFSDDSSSLTIRKVD 277
 QY 312 KEHSGRIECOG-----LDLDTMISLSEPOELLVNVSDVRYSPAAPERQSGSSLTLC 365
 Db 278 KNEAEVYVCIKENKAGQODASIHILKYFAKPK--ITYVE---NOTAMELEE--QVTLTC 328
 QY 366 EABE-----SQDLFQWLREITGOVLERGPVLQH-----DKREA 401
 Db 329 EASGDPIPSITWTSTRNISSEKASWTREKQETLDGHMVRS HARVSSLTKSIQYTD 388
 QY 402 GGGYRCYAS-----VPSIPGLNRQOLVNAIFGPPMMAFKERKVVYKEN 445
 Db 389 AGEVYCTASNTIGODSQSMYLEVOYAPKLOG-----PVAVY-----TW--EG 428
 QY 446 MVLNLSCASGHPPTISWNVNGTASEQODPQRYLSTLNLVLP--ELLE----- 495
 Db 429 NOVITCEVFAYPSPATISMPRDQLL-----PSSNYSNIKIYNTPSASYLEVTPDSEND 483
 QY 496 -GVECTASNDLGKNTSLFLELVNLTTPPDSNTTGLSTSTASPHTRANSTSTERKLE 554
 Db 484 GNTNCTAVNRIQES---LEFVLVQADTPSP---STDQVEPY---STAQVQFDE 530
 QY 555 PESRGVIVAV-----IVCIIVL-----AVLGAVLYF 581
 Db 531 PEATGGVPIILKYKAEMWAMEEYVHSHKWDAKESMEGIYIIGLKPETTYAVRLAAL-- 588
 QY 582 LYKKGKLPCKRRSGKQETLPPSRKSELVYVSKDKLPEENG 622
 Db 589 ---NGK-----GIGETISAASEFKTQPVREPSAPKLEGQMG 620

Search completed: June 28, 2002, 10:50:55
 Job time: 260 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2002, 10:45:50 ; Search time 34.31 Seconds

(without alignments)
3257.206 Million cell updates/sec

Title: US-09-653-961-2

Perfect score: 3363

Sequence: 1 MGLEPRIVCAFLAACCPCPR.....SSGDKRAPDQGEKIDLRH 646

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP archaea:.*
2: SP bacteria:.*
3: SP fungi:.*
4: SP human:.*
5: SP invertebrate:.*
6: SP mammal:.*
7: SP mhc:.*
8: SP organelle:.*
9: SP phage:.*
10: SP plant:.*
11: SP rodent:.*
12: SP virus:.*
13: SP vertebrate:.*
14: SP unclassified:.*
15: SP viirus:.*
16: SP bacteriap:.*
17: SP archaeap:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3359	99.9	646	4 Q9BRD9	Q9BRD9 homo sapien
2	3356	99.8	646	4 Q9BRD9	Q9BRD9 homo sapien
3	2642	78.6	648	11 Q9BRP1	Q9BRP1 mus musculu
4	2587	76.9	648	11 Q9BRP2	Q9BRP2 raltus norv
5	2409	71.6	606	11 Q9ES57	Q9ES57 mus musculu
6	2354	70.0	606	11 Q9ES58	Q9ES58 raltus norv
7	1104	32.8	626	13 Q9BR92	Q9BR92 gallus gall
8	1025	30.5	626	13 Q9BR80	Q9BR80 gallus gall
9	863	25.7	584	13 Q9BR91	Q9BR91 gallus gall
10	863	25.7	584	13 Q9BR92	Q9BR92 gallus gall
11	747	22.2	504	13 Q9BR93	Q9BR93 gallus gall
12	645.5	19.2	628	6 Q9WZ08	Q9WZ08 bos taurus
13	618	18.4	622	11 Q9R0B2	Q9R0B2 mus musculu
14	614	18.3	622	11 Q9R0B9	Q9R0B9 mus musculu
15	613	18.2	622	11 Q9ES55	Q9ES55 mus musculu
16	613	18.2	624	11 Q9ES56	Q9ES56 raltus norv

ALIGNMENTS

RESULT	ID	Q9BRD9	PRELIMINARY:	PRT:	646 AA.
AC	Q9BRD9	01-JUN-2001 (Tremblrel. 17, Created)			
DT	01-JUN-2001 (Tremblrel. 17, Last sequence update)				
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)				
DE	SIMILAR TO MELANOMA ADHESION MOLECULE.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RP	[1]				
RC	SEQUENCE FROM N.A.				
RA	Strausberg R.;				
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; BC006329; AAH06329.1; -				
DR	InterPro: IPR003598; Ig_C2.				
DR	InterPro: IPR003600; Ig_Like.				
DR	InterPro: IPR003006; Ig_MHC.				
DR	Pfam: PF00047; Ig_5				
DR	SMART: SM00409; IG_5				
DR	SMART: SM00410; IG_2; 3.				
KW	Immunoglobulin domain.				
SO	SEQUENCE 646 AA; 71608 MW; 1P2CG96B64B16635 CRC64;				
Query Match	99.9%; Score 3359; DB 4; Length 646;				
Best Local Similarity	99.8%; Pred. No. 1.5e-267;				
Matches 645; Conservative	1; Mismatches 0; Indels 0; Gaps 0;				
OY	1 MGLEPRIVCAFLAACCPCPRVAGVGEAEPAPPEVEVGSFLKCGLSQSGNLSHY 60				Q99K86 mus musculu
					Q35112 raltus norv
					Q9BHL3 bos taurus
					Q46634 canis faml
					Q46651 oryctolagus
					Q91B2 mus musculu
					Q19129 bos taurus
					Q96184 homo sapien
					Q965C3 homo sapien
					Q76518 caenorhabd
					Q96RW7 homo sapien
					Q96DN3 homo sapien
					Q29123 sus scrofa
					Q28939 sus scrofa
					Q96MS0 homo sapien
					Q96P7 drosophila
					Q96P7 drosophila
					Q96190 homo sapien
					Q57577 cynops pyrr
					Q94355 cynops pyrr
					Q98026 mus musculu
					Q85005 raltus norv
					Q63663 raltus norv
					Q98S4 brachydanio
					Q9VQ10 drosophila
					Q28260 canis faml
					Q967X6 drosophila
					Q9V4Y0 drosophila
					Q73633 xenopus lae
OY	61 DMFSVHKERTLIFRVQSGSGPEFEYQRLSLQDKGATLALQVTPDRIFLCGGR 120				

```
|||||
Db 61 DWFVHKEKRTLLIFRVROGQSGSEPEGEYORLSLQDGAIALJVTPODEEITFLCOGKR 120
Oy 121 PRSOEYRIQLRVYKAPPEPNIOVNPICIPVNSKEPEEVATCVGRNGYPIPOYIWKNGRP 180
Db 121 PRSOEYRIQLRVYKAPPEPNIOVNPICIPVNSKEPEEVATCVGRNGYPIPOYIWKNGRP 180
Oy 181 LKEEKNRVHIQSSQTVESGSLTYTLOSILKAQVKEKDAQFYCELYRPLPSGNHMKESRE 240
Db 181 LKEEKNRVHIQSSQTVESGSLTYTLOSILKAQVKEKDAQFYCELYRPLPSGNHMKESRE 240
Oy 241 VTVPVFPTKEVWLEVEPVGMLEKGDREVEIRCLADGNPPHFSISKQNPSTREAEEETT 300
Db 241 VTVPVFPTKEVWLEVEPVGMLEKGDREVEIRCLADGNPPHFSISKQNPSTREAEEETT 300
Oy 301 DNGVLVEPARKHSGRYECQGLDIDTMIISLSEPOELLVNVSDVRSVPAAPEROGSS 360
Db 301 DNGVLVEPARKHSGRYECQGLDIDTMIISLSEPOELLVNVSDVRSVPAAPEROGSS 360
Oy 361 LITFCEASSQDLEFQWMLREBTGOVLERGPVLOLHDLKREAGGRCVAVSPISIGLNRT 420
Db 361 LITFCEASSQDLEFQWMLREBTGOVLERGPVLOLHDLKREAGGRCVAVSPISIGLNRT 420
Oy 421 QLVNVAIFGPPMAFKERKVVYKEMVNLNLSCEASGHPRTISWNVNGTASQDODPORY 480
Db 421 QLVNVAIFGPPMAFKERKVVYKEMVNLNLSCEASGHPRTISWNVNGTASQDODPORY 480
Oy 481 LSTLVNLVLPBELLEGTVECTASNDLGKNTSILFLELVNLTTLTPDSNTTGLSTASPH 540
Db 481 LSTLVNLVLPBELLEGTVECTASNDLGKNTSILFLELVNLTTLTPDSNTTGLSTASPH 540
Oy 541 TRANSTSTERKLPPEPSRGVIVAVICILVAVLGAVILEYFKKKKLCRCRSGQOETL 600
Db 541 TRANSTSTERKLPPEPSRGVIVAVICILVAVLGAVILEYFKKKKLCRCRSGQOETL 600
Oy 601 PPSRKSSELYVEVKSOKLPEEMGLQSSGDKRAPDQGEKYIDLKH 646
Db 601 PPSRKSSELYVEVKSOKLPEEMGLQSSGDKRAPDQGEKYIDLKH 646

RESULT 2
Oy 095812 PRELIMINARY: PRT: 646 AA.
AC 095812;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DE CELL SURFACE GLYCOPROTEIN P1H12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Gul L., Chang L., Browne P.V., Heibel R.P.;
RT "P1H12 from human umbilical vein endothelial cells."
RL EMBL: AF089868; AAD17799.1;
DR InterPro: IPR003598; Iq_c2.
DR InterPro: IPR003600; Iq_1like.
DR InterPro: IPR003006; Iq_MHC.
DR Pfam: PF00047; Iq: 5.
DR SMART: SM00408; IqC2: 2.
DR SMART: SM00410; Iq_1like: 2.
KW Immunoglobulin domain.
SQ SEQUENCE 646 AA: 71665 MW: 185FB8AD930738E CRC64:
```

```
|||||
Db 61 DWFVHKEKRTLLIFRVROGQSGSEPEGEYORLSLQDGAIALJVTPODEEITFLCOGKR 120
Oy 61 DWFVHKEKRTLLIFRVROGQSGSEPEGEYORLSLQDGAIALJVTPODEEITFLCOGKR 120
Db 61 DWFVHKEKRTLLIFRVROGQSGSEPEGEYORLSLQDGAIALJVTPODEEITFLCOGKR 120
Oy 121 PRSOEYRIQLRVYKAPPEPNIOVNPICIPVNSKEPEEVATCVGRNGYPIPOYIWKNGRP 180
Db 121 PRSOEYRIQLRVYKAPPEPNIOVNPICIPVNSKEPEEVATCVGRNGYPIPOYIWKNGRP 180
Oy 181 LKEEKNRVHIQSSQTVESGSLTYTLOSILKAQVKEKDAQFYCELYRPLPSGNHMKESRE 240
Db 181 LKEEKNRVHIQSSQTVESGSLTYTLOSILKAQVKEKDAQFYCELYRPLPSGNHMKESRE 240
Oy 241 VTVPVFPTKEVWLEVEPVGMLEKGDREVEIRCLADGNPPHFSISKQNPSTREAEEETT 300
Db 241 VTVPVFPTKEVWLEVEPVGMLEKGDREVEIRCLADGNPPHFSISKQNPSTREAEEETT 300
Oy 301 DNGVLVEPARKHSGRYECQGLDIDTMIISLSEPOELLVNVSDVRSVPAAPEROGSS 360
Db 301 DNGVLVEPARKHSGRYECQGLDIDTMIISLSEPOELLVNVSDVRSVPAAPEROGSS 360
Oy 361 LITFCEASSQDLEFQWMLREBTGOVLERGPVLOLHDLKREAGGRCVAVSPISIGLNRT 420
Db 361 LITFCEASSQDLEFQWMLREBTGOVLERGPVLOLHDLKREAGGRCVAVSPISIGLNRT 420
Oy 421 QLVNVAIFGPPMAFKERKVVYKEMVNLNLSCEASGHPRTISWNVNGTASQDODPORY 480
Db 421 QLVNVAIFGPPMAFKERKVVYKEMVNLNLSCEASGHPRTISWNVNGTASQDODPORY 480
Oy 481 LSTLVNLVLPBELLEGTVECTASNDLGKNTSILFLELVNLTTLTPDSNTTGLSTASPH 540
Db 481 LSTLVNLVLPBELLEGTVECTASNDLGKNTSILFLELVNLTTLTPDSNTTGLSTASPH 540
Oy 541 TRANSTSTERKLPPEPSRGVIVAVICILVAVLGAVILEYFKKKKLCRCRSGQOETL 600
Db 541 TRANSTSTERKLPPEPSRGVIVAVICILVAVLGAVILEYFKKKKLCRCRSGQOETL 600
Oy 601 PPSRKSSELYVEVKSOKLPEEMGLQSSGDKRAPDQGEKYIDLKH 646
Db 601 PPSRKSSELYVEVKSOKLPEEMGLQSSGDKRAPDQGEKYIDLKH 646

RESULT 3
Oy 09EP1 PRELIMINARY: PRT: 648 AA.
AC 09EP1;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE L-GICERIN/MUC18.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Taira E., Okumura S., Miki N.;
RT "mouse gicerin/MUC18."
RL Submitted (DEC-1993) to the EMBL/Genbank/DBJ databases.
DR EMBL: AB035508; BAB16050.1;
DR MGI: MGI:193396; Mcam.
DR InterPro: IPR003599; Iq.
DR InterPro: IPR003598; Iq_c2.
DR InterPro: IPR003600; Iq_1like.
DR InterPro: IPR003006; Iq_MHC.
DR Pfam: PF00047; Iq: 5.
DR SMART: SM00409; Iq: 4.
DR SMART: SM00408; IqC2: 4.
DR SMART: SM00410; Iq_1like: 5.
```


KW Immunoglobulin domain.
SQ SEQUENCE 648 AA: 71511 MW: EA863AB35B8C28FD CRC64:

Query Match 78.6%; Score 2642; DB 11; Length 648;
Best Local Similarity 76.4%; Pred. No. 1.4e-208;
Matches 495; Conservative 73; Mismatches 78; Indels 2; Gaps 1;

1 MGPRVCAFLAACCCPRVAVPGEAEQ--PAPELVEVVGSTALLKCGISQSGNLS 58
1 MGPRVCAFLAACCCPRVAVPGEAEQ--PAPELVEVVGSTALLKCGISQSGNLS 60
1 MGPRVCAFLAACCCPRVAVPGEAEQ--PAPELVEVVGSTALLKCGISQSGNLS 60
59 HDWFSVHKERTLIFRVQGGOGSEPEYBQRLSLDGRATLALQVTPDERIFLCOG 118
59 HDWFSVHKERTLIFRVQGGOGSEPEYBQRLSLDGRATLALQVTPDERIFLCOG 120
61 QVWFLIHKEROILIFRVHGGKGGSEPEYBQRLSLDGRATLALQVTPDERIFLCOG 120
119 KPRSGEYRIOLRVYKAPPEPNIQVNPICIPYNSKEPEEVATCGRNKGPITQVYIWKNG 178
119 KPRSGEYRIOLRVYKAPPEPNIQVNPICIPYNSKEPEEVATCGRNKGPITQVYIWKNG 180
121 KPRSGEYRIOLRVYKAPPEPNIQVNPICIPYNSKEPEEVATCGRNKGPITQVYIWKNG 180
121 KPRSGEYRIOLRVYKAPPEPNIQVNPICIPYNSKEPEEVATCGRNKGPITQVYIWKNG 238
179 RPLKEKNRVHIOSSQTVSSGLTYLQSLIKALQVKEKDAQFYCELNYRLPSGNHMKES 238
179 RPLKEKNRVHIOSSQTVSSGLTYLQSLIKALQVKEKDAQFYCELNYRLPSGNHMKES 240
181 LPLQEEENRVHIOSSQTVSSGLTYLQSLIKALQVKEKDAQFYCELNYRLPSGNHMKES 240
239 REVTVVFPYFTEKRWLEVEPVGMKKGDRVEIRCLADGNPPHPSISKONPSTREAEET 298
239 REVTVVFPYFTEKRWLEVEPVGMKKGDRVEIRCLADGNPPHPSISKONPSTREAEET 300
241 KEVTVVFPYFTEKRWLEVEPVGMKKGDRVEIRCLADGNPPHPSISKONPSTREAEET 300
299 TNDNGVYVLEPARKHSGRKECOGLDPTMISLSEPOELLVNVSVSPVSPAPEROEG 358
299 TNDNGVYVLEPARKHSGRKECOGLDPTMISLSEPOELLVNVSVSPVSPAPEROEG 360
301 TDENGILSLPEPAKHSGILYOCOSLDEFTTSSDPLELLVNVSVSPVSPAPEROEG 360
359 SSVLTTCFAESSODLEFOWLREETGOVLENGPVLQDLHDKREGAGGYRCVAVSPSTPGIN 418
359 SSVLTTCFAESSODLEFOWLREETGOVLENGPVLQDLHDKREGAGGYRCVAVSPSTPGIN 420
361 ESLVLTCAESNODLEFOWLREETGOVLENGPVLQDLHDKREGAGGYRCVAVSPSTPGIN 420
419 RTQLVNVALFGPMAAKERKRWKENVNLNLSCEASGHPPTISMVNGTASQDODPO 478
419 RTQLVNVALFGPMAAKERKRWKENVNLNLSCEASGHPPTISMVNGTASQDODPO 480
421 RTQLVNVALFGPMAAKERKRWKENVNLNLSCEASGHPPTISMVNGTASQDODPO 480
479 RVLSLTVNLYVPELLETGECTASNDLGKNTSILFLELVNLTPTTSPDNTTGTSTAS 538
479 RVLSLTVNLYVPELLETGECTASNDLGKNTSILFLELVNLTPTTSPDNTTGTSTAS 540
481 TVVSTLVNLYVPELLETGECTASNDLGKNTSILFLELVNLTPTTSPDNTTGTSTAS 540
539 PHTRANSTSTERKLEPEESRGVIVAVYICILVAVLAVGAVLYELKKGKLPGRSGKOEI 598
539 PHTRANSTSTERKLEPEESRGVIVAVYICILVAVLAVGAVLYELKKGKLPGRSGKOEI 600
541 PHTRANSTSTERKLEPEESRGVIVAVYICILVAVLAVGAVLYELKKGKLPGRSGKOEI 600
599 TLPSRKSELYVEVSKDKLPEEMGLLOSGSGDKRAPGOGKXYIDLKH 646
599 TLPSRKSELYVEVSKDKLPEEMGLLOSGSGDKRAPGOGKXYIDLKH 648
601 TLPSRKSELYVEVSKDKLPEEMGLLOSGSGDKRAPGOGKXYIDLKH 648

RESULT 4
PRELIMINARY: PRT; 648 AA.

AC 09EPF2; 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE L-GICERIN/MUC18.
GN L-GICERIN/MUC18.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RA Tajima E., Okumura S., Miki N.;
RT "Functional analysis of rat gicerin/MUC18.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
EMBL: AB035506; BAB16048.1;

DR InterPro: IPR003599; I9.
DR InterPro: IPR003598; I9_C2.
DR InterPro: IPR003600; I9_Like.
DR InterPro: IPR003606; I9_MHC.
DR Pfam: PF00047; I9; 5.
DR SMART: SM00409; I9; 4.
DR SMART: SM00408; I9C2; 3.
DR SMART: SM00410; I9_Like; 5.
KW Immunoglobulin domain.
SQ SEQUENCE 648 AA: 71372 MW: F8C334E0F1938A9F CRC64:

Query Match 76.9%; Score 2587; DB 11; Length 648;
Best Local Similarity 74.5%; Pred. No. 4.8e-204;
Matches 483; Conservative 80; Mismatches 83; Indels 2; Gaps 1;

1 MGPRVCAFLAACCCPRVAVPGEAEQ--PAPELVEVVGSTALLKCGISQSGNLS 58
1 MGPRVCAFLAACCCPRVAVPGEAEQ--PAPELVEVVGSTALLKCGISQSGNLS 60
1 MGPRVCAFLAACCCPRVAVPGEAEQ--PAPELVEVVGSTALLKCGISQSGNLS 60
59 HDWFSVHKERTLIFRVQGGOGSEPEYBQRLSLDGRATLALQVTPDERIFLCOG 118
59 HDWFSVHKERTLIFRVQGGOGSEPEYBQRLSLDGRATLALQVTPDERIFLCOG 120
61 QVWFLIHKEROILIFRVHGGKGGSEPEYBQRLSLDGRATLALQVTPDERIFLCOG 120
119 KPRSGEYRIOLRVYKAPPEPNIQVNPICIPYNSKEPEEVATCGRNKGPITQVYIWKNG 178
119 KPRSGEYRIOLRVYKAPPEPNIQVNPICIPYNSKEPEEVATCGRNKGPITQVYIWKNG 180
121 KPRSGEYRIOLRVYKAPPEPNIQVNPICIPYNSKEPEEVATCGRNKGPITQVYIWKNG 180
121 KPRSGEYRIOLRVYKAPPEPNIQVNPICIPYNSKEPEEVATCGRNKGPITQVYIWKNG 238
179 RPLKEKNRVHIOSSQTVSSGLTYLQSLIKALQVKEKDAQFYCELNYRLPSGNHMKES 238
179 RPLKEKNRVHIOSSQTVSSGLTYLQSLIKALQVKEKDAQFYCELNYRLPSGNHMKES 240
181 RPLKEKNRVHIOSSQTVSSGLTYLQSLIKALQVKEKDAQFYCELNYRLPSGNHMKES 240
239 REVTVVFPYFTEKRWLEVEPVGMKKGDRVEIRCLADGNPPHPSISKONPSTREAEET 298
239 REVTVVFPYFTEKRWLEVEPVGMKKGDRVEIRCLADGNPPHPSISKONPSTREAEET 300
241 KEVTVVFPYFTEKRWLEVEPVGMKKGDRVEIRCLADGNPPHPSISKONPSTREAEET 300
299 TNDNGVYVLEPARKHSGRKECOGLDPTMISLSEPOELLVNVSVSPVSPAPEROEG 358
299 TNDNGVYVLEPARKHSGRKECOGLDPTMISLSEPOELLVNVSVSPVSPAPEROEG 360
301 TDENGILSLPEPAKHSGILYOCOSLDEFTTSSDPLELLVNVSVSPVSPAPEROEG 360
359 SSVLTTCFAESSODLEFOWLREETGOVLENGPVLQDLHDKREGAGGYRCVAVSPSTPGIN 418
359 SSVLTTCFAESSODLEFOWLREETGOVLENGPVLQDLHDKREGAGGYRCVAVSPSTPGIN 420
361 DSVLTTCFAESSODLEFOWLREETGOVLENGPVLQDLHDKREGAGGYRCVAVSPSTPGIN 420
419 RTQLVNVALFGPMAAKERKRWKENVNLNLSCEASGHPPTISMVNGTASQDODPO 478
419 RTQLVNVALFGPMAAKERKRWKENVNLNLSCEASGHPPTISMVNGTASQDODPO 480
421 RTRRVSVGIFGSPMAAKERKRWKENVNLNLSCEASGHPPTISMVNGTASQDODPO 480
479 RVLSLTVNLYVPELLETGECTASNDLGKNTSILFLELVNLTPTTSPDNTTGTSTAS 538
479 RVLSLTVNLYVPELLETGECTASNDLGKNTSILFLELVNLTPTTSPDNTTGTSTAS 540
481 TVVSTLVNLYVPELLETGECTASNDLGKNTSILFLELVNLTPTTSPDNTTGTSTAS 540
539 PHTRANSTSTERKLEPEESRGVIVAVYICILVAVLAVGAVLYELKKGKLPGRSGKOEI 598
539 PHTRANSTSTERKLEPEESRGVIVAVYICILVAVLAVGAVLYELKKGKLPGRSGKOEI 600
541 PHTRANSTSTERKLEPEESRGVIVAVYICILVAVLAVGAVLYELKKGKLPGRSGKOEI 600
599 TLPSRKSELYVEVSKDKLPEEMGLLOSGSGDKRAPGOGKXYIDLKH 646
599 TLPSRKSELYVEVSKDKLPEEMGLLOSGSGDKRAPGOGKXYIDLKH 648
601 TLPSRKSELYVEVSKDKLPEEMGLLOSGSGDKRAPGOGKXYIDLKH 648

RESULT 5
PRELIMINARY: PRT; 606 AA.

AC 09ESS7; 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE S-GICERIN/MUC18.
GN MAM OR S-GICERIN/MUC18.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BA1B/C;
RA Taira E., Okumura S., Miki N.;
RT "mouse s-glycerin/MUC18."
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB035509; BAB16051.1;
DR MGI; MGI:1933966; Mcam.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003600; Ig_like.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig_5.
DR SMART; SM00409; Ig; 4.
DR SMART; SM00408; IgC2; 4.
DR SMART; SM00410; Ig_Like; 5.
KW Immunoglobulin domain.
SQ SEQUENCE 606 AA; 66867 MW; 8F3CBAC0A5F29B10 CRC64;

Query Match 71.6%; Score 2409; DB 11; Length 606;
Best Local Similarity 75.0%; Pred. No. 1,9e-189;
Matches 450; Conservative 72; Mismatches 76; Indels 2; Gaps 1;

QY 1 MGLPRLVCAFLAACCCCPVAVPGEAEO--PAPELVEVEVSTALLKGLSOGNLS 58
DB 1 MGLPRLVCAFLAACCCCPVAVPGEAEO--PAPELVEVEVSTALLKGLSOGNLS 58
QY 59 HDMFVSVHKERKRLIFRVGQSGSEPEGEORLSLODAGATLTOYTPODESTIFLCOG 118
DB 61 QVWFILHKEQRLIFRVHGQSGSEPEGEORLSLODAGATLTOYTPODESTIFLCOG 120
QY 119 KRPSOEYRIOLRVYKAPEEPNIQVNPICIPYNSKEPEEVAATCVGRNGYPIPOYIWKNG 178
DB 121 KRRLODHYVLEQVAKAPEEPTIQANVGVINVDROELREVAATCVGRNGYPIPOYIWKNG 180
QY 179 RPLKEKNRVHIQSSQVTESSGLYTLQSLKLAQLYKEDKQAFCELYNLPYSGNHKES 238
DB 181 RPLQEBENRVHIQSSQVTESSGLYTLQSLKLAQLYKEDKQAFCELYNLPYSGNHKES 240
QY 239 REVTVAVFPTKRVWLEVEPVGMKKEGDRVETICRLADGNPPHPSISKNPSTREAEEET 298
DB 241 KEVTVAVFPTKRVWLEVEPVGMKKEGDRVETICRLADGNPPHPSISKNPSTREAEEET 300
QY 299 TNDNGVLYLEPAKKEHSGRYECQGLDITMTISLSEPOELLVNVYSDVRYSPAAPROEG 358
DB 301 TNDNGVLYLEPAKKEHSGRYECQGLDITMTISLSEPOELLVNVYSDVRYSPAAPROEG 360
QY 359 SSLTITCEAESODLEPQMLREENGQVLENGPVYLQHLDKREAGGGRVAVSPISPGIN 418
DB 361 SSLTITCEAESODLEPQMLREENGQVLENGPVYLQHLDKREAGGGRVAVSPISPGIN 420
QY 419 RTQLVNVAIFGPPMAKEREKRVKEMVNLNLSCEASGHPRPITSNVNGTSEDDOPQ 478
DB 421 RTQLVNVAIFGPPMAKEREKRVKEMVNLNLSCEASGHPRPITSNVNGTSEDDOPQ 480
QY 479 RVLSTLVNLYTPLELLENGVECTASNDLGKNTSLFLDELVNLTTLPDSTNTGLSTSTAS 538
DB 481 TVVSTLVNLYTPLELLENGVECTASNDLGKNTSLFLDELVNLTTLPDSTNTGLSTSTAS 540
QY 539 PHTRANSTSTERKLPPEPSRGVIVAVIYCIYLAVALGAVLELKKKRLCRSGKOET 598
DB 541 PHTRANSTSTERKLPPEPSRGVIVAVIYCIYLAVALGAVLELKKKRLCRSGKOET 600

RESULT 6
Q9ESS8 PRELIMINARY; PRT; 606 AA.
AC Q9ESS8;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE S-GICERIN/MUC18.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=HEART;
RA Taira E., Okumura S., Miki N.;
RT "rat s-glycerin/MUC18."
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB035507; BAB16049.1;
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003600; Ig_Like.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig_5.
DR SMART; SM00409; Ig; 4.
DR SMART; SM00408; IgC2; 3.
DR SMART; SM00410; Ig_Like; 5.
KW Immunoglobulin domain.
SQ SEQUENCE 606 AA; 66727 MW; 6D0892443E8B8C47 CRC64;

Query Match 70.0%; Score 2354; DB 11; Length 606;
Best Local Similarity 73.0%; Pred. No. 6.3e-185;
Matches 438; Conservative 79; Mismatches 81; Indels 2; Gaps 1;

QY 1 MGLPRLVCAFLAACCCCPVAVPGEAEO--PAPELVEVEVSTALLKGLSOGNLS 58
DB 1 MGLPRLVCAFLAACCCCPVAVPGEAEO--PAPELVEVEVSTALLKGLSOGNLS 58
QY 59 HDMFVSVHKERKRLIFRVGQSGSEPEGEORLSLODAGATLTOYTPODESTIFLCOG 118
DB 61 QVWFILHKEQRLIFRVHGQSGSEPEGEORLSLODAGATLTOYTPODESTIFLCOG 120
QY 119 KRPSOEYRIOLRVYKAPEEPNIQVNPICIPYNSKEPEEVAATCVGRNGYPIPOYIWKNG 178
DB 121 KRRLODHYVLEQVAKAPEEPTIQANVGVINVDROELREVAATCVGRNGYPIPOYIWKNG 180
QY 179 RPLKEKNRVHIQSSQVTESSGLYTLQSLKLAQLYKEDKQAFCELYNLPYSGNHKES 238
DB 181 RPLQEBENRVHIQSSQVTESSGLYTLQSLKLAQLYKEDKQAFCELYNLPYSGNHKES 240
QY 239 REVTVAVFPTKRVWLEVEPVGMKKEGDRVETICRLADGNPPHPSISKNPSTREAEEET 298
DB 241 KEVTVAVFPTKRVWLEVEPVGMKKEGDRVETICRLADGNPPHPSISKNPSTREAEEET 300
QY 299 TNDNGVLYLEPAKKEHSGRYECQGLDITMTISLSEPOELLVNVYSDVRYSPAAPROEG 358
DB 301 TNDNGVLYLEPAKKEHSGRYECQGLDITMTISLSEPOELLVNVYSDVRYSPAAPROEG 360
QY 359 SSLTITCEAESODLEPQMLREENGQVLENGPVYLQHLDKREAGGGRVAVSPISPGIN 418
DB 361 SSLTITCEAESODLEPQMLREENGQVLENGPVYLQHLDKREAGGGRVAVSPISPGIN 420
QY 419 RTQLVNVAIFGPPMAKEREKRVKEMVNLNLSCEASGHPRPITSNVNGTSEDDOPQ 478
DB 421 RTQLVNVAIFGPPMAKEREKRVKEMVNLNLSCEASGHPRPITSNVNGTSEDDOPQ 480
QY 479 RVLSTLVNLYTPLELLENGVECTASNDLGKNTSLFLDELVNLTTLPDSTNTGLSTSTAS 538
DB 481 TVVSTLVNLYTPLELLENGVECTASNDLGKNTSLFLDELVNLTTLPDSTNTGLSTSTAS 540
QY 539 PHTRANSTSTERKLPPEPSRGVIVAVIYCIYLAVALGAVLELKKKRLCRSGKOET 598
DB 541 PHTRANSTSTERKLPPEPSRGVIVAVIYCIYLAVALGAVLELKKKRLCRSGKOET 600

RESULT 7
Q98922 PRELIMINARY; PRT; 626 AA.
ID Q98922

AC 098922: 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE HEMCAM PRECURSOR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H.B190V+; TISSUE=THYMUS;
RA MEDLINE=97133433; PubMed=8978830;
VA Valino O., Dunon D., Aissi F., Dangy J.P., McNaghy K.M., Imhof B.A.;
RT "HEMCAM, an adhesion molecule expressed by c-kit+ progenitors.";
RL J. Cell Biol. 135:1655-1668(1996).
DR EMBL: Y08855; CA70080.1; -
DR InterPro: IPR003598; Ig_C2.
DR InterPro: IPR003600; Ig_Like.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig_5.
DR SMART: SM00408; IgC2; 1.
DR SMART: SM00410; Ig_Like; 3.
KW Immunoglobulin domain; Signal.
FT SIGNAL 1 35 POTENTIAL.
FT CHAIN 36 626 POTENTIAL.
SQ SEQUENCE 626 AA; 69104 MW; 27F8F2F47044E163 CRC64;

Query Match 32.8%; Score 1104; DB 13; Length 626;
Best Local Similarity 36.3%; Pred. No. 4.3e-82;
Matches 251; Conservative 116; Mismatches 221; Indels 68; Gaps 16;

QY 8 CAFILAACCCPRVAVGEAEOPAPELVEVEGSTALLKCGLS-QSGNGNSHYDMSVH 66
DB 22 CPELCCLLCC-----GAAGRLVEYMPAVLEVEIGSTABLECSFISPGNASTSIEMFYVN 76
QY 67 KEKRTLFRVRGOG-----QSEPEYBQRLSDRGATLALQVTPDE-RIFLCQ 117
DB 77 RRP-----LRSGAVRHHAGVRIDEYEYSRLSVGDKA-LSISKVTRQDNARTICQ 129
QY 118 --GKRPRSGEYRIOLRYKAPBEENIOVNPGLGIPVNSKEPEVATCGVGRNCPYIOWY 175
DB 130 VGADSGVGESRTELYTKIPAPPEITPNSAGIPAOQNDMKIAQCTSENSFSPNTTWY 189
QY 176 KNGRPLEKKNRVHIQSSQVESSGLYTLQSLKAOLYKEDKDAQFCENLYRPLPSNMH 235
DB 190 KNGPELQEDKTKILTLTVRESNGLYTVSTLFSKYTREDRNSLFCHTYHYWLOGQMT 249
QY 236 KESREYVVPVYPTPEKWLVEEP-VGMKEDDVEIRCLADGNPPPHFSISKQ--NPSTR 292
DB 250 KDSRYVNTVYFPTPEHVELRVATNAGIVKEGDDVKLVCDADGNPAVPSFRRRLGDSWQ 309
QY 293 EAESETTNDNCVLEVPARKHSGREYEGOGIDLTMTSLSEPOELLVNVYSVRY--SP 350
DB 310 DMTSLATNDGVLMLHNVSKSSGLYRCQTLDDMTQHEGD-VELVYNTIEGVYKMP 368
QY 351 AAPERQSGSSLTLLTCEASSODLEFOWLREFTGVLERGPVLOLHDLKRRAGGGRCVAS 410
DB 369 SSP-LHGGDSVRLSCTASHSPVKLDYQW-RDARGKVAEGNQLLTNTLFTTSSNFSGRVK 426
QY 411 VPSITGLNRQOLVNVALFEGPPMAFFKRWVYKENVMLNLSCEASGHPRTTSMVNWGTA 470
DB 427 ARSPVLEQSSQVAVVAKGKPRIVASAPPLYVRQDEVINTCKAIAFPQPSFMSINGTT 486
QY 471 SEQDDQPRVLTNLVYTPELLEETGECTASNDLGKNTSILFLELVNLTLLPDSITT 530
DB 487 HEY-MENQHNASNLTVAVSHDLRAGAMCYSNALGVSE-----KHLQLDQKPS----- 535
QY 531 GLSTSTASPHTRANSTSTERKLPEPESRGVYVAVYICILVAVLGAVALYFLKKKKLPC 590
DB 536 -----ESKGIITVIAIIVCIYVAVLGSIIYFLKKKKKISC 570

QY 591 RRSKGQETLTPSSRKSSELYVEVKSCLKPEEMGLQSSGDKRARGQGEKYIDLKH 646
DB 571 GRSGQDITPEARKDNVVEKSDKLSEBAGILLQANGKRRSPADOSEKYIDLKN 626

RESULT 8
QY 090880 PRELIMINARY; PRT; 626 AA.
AC 090880:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE L-GICERIN.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LEGHORN; TISSUE=GIZARD.
RA MEDLINE=96081930; PubMed=749368;
VA Taira E., Nagino T., Taniura H., Takaha N., Kim C., Kuo C., Li B.,
RA Higuchi H., Miki N.;
RT "Expression and functional analysis of a novel isoform of glicerin, an
immunoglobulin superfamily cell adhesion molecule.";
RL J. Biol. Chem. 270:28681-28687(1995).
DR EMBL: D49849; BAA08648.1; -
DR InterPro: IPR003598; Ig_C2.
DR InterPro: IPR003600; Ig_Like.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig_5.
DR SMART: SM00408; IgC2; 1.
DR SMART: SM00410; Ig_Like; 2.
KW Immunoglobulin domain.
SQ SEQUENCE 626 AA; 69074 MW; C4791EBC2EC55133 CRC64;

Query Match 30.5%; Score 1025; DB 13; Length 626;
Best Local Similarity 37.4%; Pred. No. 1.3e-75;
Matches 248; Conservative 114; Mismatches 219; Indels 82; Gaps 19;

QY 8 CAFILAACCCPRVAVGEAEOPAPELVEVEGSTALLKCGLS-QSGNGNSHYDMSVH 62
DB 22 CPELCCLLCC-----GAAGR-----EYVMPSSA-----GSGRHSQAGVOLLH-PW 63
QY 63 FSVHKEKRLIFR--VRGOG-----QSEPEYBQRLSDRGATLALQVTPDE 111
DB 64 ECLLHLRVVLCQPRPLRSGEAVRHHAGVRIDEYEYSRLSVGDKA-LSISKVTRQDN 122
QY 112 -RIFLCQ--GKRPRSGEYRIOLRYKAPBEENIOVNPGLGIPVNSKEPEVATCGVGRNCPY 168
DB 123 ARFTIQVGADSGVGESRTELYTKIPAPPEITPNSAGIPAOQNDMKIAQCTSENSF 182
QY 169 IPQVLYKNGRPLEKKNRVHIQSSQVESSGLYTLQSLKAOLYKEDKDAQFCENLYR 228
DB 183 SPVITWYKNGPELQEDKTKILTLTVRESNGLYTVSTLFSKYTREDRNSLFCHTYHYW 242
QY 229 LPSGNHMKSSREYTPVFPPTPEKWLVEEP-VGMKEDDVEIRCLADGNPPPHFSISKQ 287
DB 243 LOGQMTKDSPRVNTVYFPTPEHVELRVATNAGIVKEGDDVKLVCDADGNPAVPSFRR 302
QY 288 --NPSTREAEETTNDNCVLEVPARKHSGREYEGOGIDLTMTSLSEPOELLVNVYS 345
DB 303 ELGDSQDWTSLADNRDGVLMHNVSKSSGLYRCQTLDDMTQHEGD-VELVYNTIEG 361
QY 346 VRY--SPAAPERQSGSSLTLLTCEASSODLEFOWLREFTGVLERGPVLOLHDLKRRAGG 403
DB 362 VQVKMEPSSP-LHGGDSVRLSCTASHSPVKLDYQW-RDARGKVAEGNQLLTNTLFTTSS 419
QY 404 GYRCAVASPSTIGLNRQOLVNVALFEGPPMAFFKRWVYKENVMLNLSCEASGHPRTTIS 463

Db 420 NESCVRKARSVPGLEQSKOYAAVAVKGRPRIVATISAPLYRQDEVINLTCKAIAFPQPSFH 479
 QY 464 MNVNGTASSEDODPQRLVSLNVLVTPLELTGETECTASNDLCKNSIFLELVNLTTLT 523
 Db 480 WSVNGTTHVY-MENQHMASNLTVRVSHDLRAGAMCRVSNALGVSE----KHQLDQK 533
 QY 524 PDSVTTTGLSTSPHTRANSTSTERKLPESRGVIVAVICILVAVLAVLYFLY 583
 Db 534 PS-----ESKGIIVAVICILVAVLAVLSIIFLH 563
 QY 584 KKGALPCRSKQEBITLPPSRKSELVYVKSDDLPEEMGLLGSSGDKRAPDQGEKYID 643
 Db 564 KKGATSCGRSGKODITKPEARKDKNVVEKSDKLSEAGLLOGANAEKRSPADQSEKYID 623
 QY 644 LRH 646
 Db 624 LRN 626

RESULT 9
 Q98921 PRELIMINARY; PRT: 584 AA.
 ID Q98921
 AC Q98921:
 DT 01-FEB-1997 (TREMBlrel. 02, Created)
 DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE HEMCAM PRECURSOR.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H.B190V+; TISSUE=THYMUS;
 RA MEDLINE=9713433; PubMed=8978830;
 RX Valiño O., Dunon D., Alsai F., Dangy J.P., McNaghy K.M., Imhof B.A.;
 RT "HEMCM, an adhesion molecule expressed by c-kit+ progenitors.";
 RL J. Cell Biol. 135:1655-1668(1996).
 DR EMBL: Y08856; CA70081.1; -
 DR InterPro: IPR003598; Iq_c2.
 DR InterPro: IPR003600; Iq_like.
 DR InterPro: IPR003006; Iq_MHC.
 DR Pfam: PF00047; Iq; 5.
 DR SMART: SM00408; IqC2; 1.
 DR SMART: SM00410; Iq_like; 3.
 KM Immunoglobulin domain; Signal.
 FT SIGNAL 1 35
 CHAIN 36 584 POTENTIAL.
 FT SIGNAL 1 35
 SQ SEQUENCE 584 AA; 64422 MW; 5972D946604AF6F3 CRC64;

Query Match 27.8%; Score 936; DB 13; Length 584;
 Best Local Similarity 35.9%; Pred. No. 2, 5e-66;
 Matches 218; Conservative 109; Mismatches 213; Indels 68; Gaps 16;

QY 8 CAFLAACCCECCRVAGVGEAQPAPLEVEVSGTALCKGLS-QSOGNLSHWDFSVH 66
 Db 22 CFFLLCLLCC-----GAAGRLEVMYPAVLEVEIGSTALECSFSLPGNASFTSIEMFVN 76
 QY 67 KKRRLIFRVVROGQ-----QSEPEYEQRLSLDRGATLAIQVTPODE-RIFLQ 117
 Db 77 RRP-----LRSGEAVRHNASGVRIDETEYSRLSVGEDKA-LTISKVTRQDNARTFICQ 129
 QY 118 --GKRPRSEYRIQLRVYKAPPEEPINQVPIGIPVNSKEPEEVAICVGRNGYPIPIQVLY 175
 Db 130 VGADSGGVESRTELYTKIPAPPETPNSAGIPQASNDMLKIACTSENSPSPNTIY 189
 QY 176 KNGRLKEEKNRVHIOSSQTVESGLYTLQSLTKALQVLEKDAQFYCELNRLPSGNHM 235
 Db 190 KNGEPLQGEDEKTKITITTVLRNSNGIYTVSLFSKYTREDNSLFLCHVHWLQGGQKRT 249
 QY 236 KESREVTVPVYPTKEVWLEVP-VGMLEKGRVETIKRLADGNPPPHFSISKQ--NPSTR 292

Db 250 KDSPRVNTVFEPTREHVELRVATNAGIVEGDDVLTVDADGNPAVPSPFRRELGSWQ 309
 QY 293 EMBEETDNGVLYLEPAKREHSGRYEGGLDLDWISLSLSPPELVNYSVDRV--SP 350
 Db 310 DWFTSLADTNDGVLMHNHNSKSSSGLRCOTLDLDMTQHEGD-VELVYNIIEGVQVKNRP 368
 QY 351 AAPEROESLTLTCEAESODLEFOWLREETGOVLERGPVLQHLDKREAGGGRVAVS 410
 Db 369 SSP-LHEGDSVRLSCGTAHSPVKLDYQW--RDARGRVAREGNQLLTNLTNFTSISNFCVVK 426
 QY 411 VSIPELNTQOLVNAVIFPPMAKREKRYVYKEMVNLNLSGASGHRPITSNVNCTA 470
 Db 427 ARSVPELEQSKOYAAVAVKGRPRIVATISAPLYRQDEVINLTCKAIAFPQPSFH 486
 QY 471 SEDODDQRLVSLNVLVTPLELTGETECTASNDLCKNSIFLELVNLTTLTPDSNTTT 530
 Db 487 HEY-MENQHMASNLTVRVSHDLRAGAMCRVSNALGVSE----KHQLDQKPS----- 535
 QY 531 GLSTSPASPHTRANSTSTERKLPESRGVIVAVICILVAVLAVLYFLYKKGKLP 590
 Db 536 -----ESKGIIVAVICILVAVLAVLSIIFLHKKKATSC 570
 QY 591 RRSKQEI 598
 Db 571 GRSKQDI 578

RESULT 10
 Q90989 PRELIMINARY; PRT: 584 AA.
 ID Q90989
 AC Q90989:
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE S-GICERIN.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LEGHORN; TISSUE=GIZZARD;
 RA MEDLINE=94213753; PubMed=8161457;
 RX Taira E., Takaha N., Taniura H., Kim C., Miki N.;
 RT "Molecular cloning and functional expression of glicerol, a novel cell
 RT adhesion molecule that binds to neurite outgrowth factor.";
 RL Neuron 12:861-872(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LEGHORN; TISSUE=GIZZARD;
 RX MEDLINE=96081930; PubMed=749388;
 RA Taira E., Nagino T., Taniura H., Takaha N., Kim C., Kuo C., Li B.,
 RA Higuchi H., Miki N.;
 RT "Expression and functional analysis of a novel isoform of glicerol, an
 RT immunoglobulin superfamily cell adhesion molecule.";
 RL J. Biol. Chem. 270:28681-28687(1995).
 DR EMBL: D38559; BAA07563.1; -
 DR InterPro: IPR003598; Iq_c2.
 DR InterPro: IPR003600; Iq_like.
 DR InterPro: IPR003006; Iq_MHC.
 DR Pfam: PF00047; Iq; 5.
 DR SMART: SM00408; IqC2; 1.
 DR SMART: SM00410; Iq_like; 2.
 KM Immunoglobulin domain.
 SQ SEQUENCE 584 AA; 64378 MW; 876EC3E920BA92F CRC64;

Query Match 25.7%; Score 863; DB 13; Length 584;
 Best Local Similarity 35.1%; Pred. No. 2, 6e-62;
 Matches 216; Conservative 107; Mismatches 210; Indels 82; Gaps 19;

```
QY 8 CAFLAACCCPRVAGVGEAEOPAPELVEVNGSTALLKCGI-----SOSQGNLSHVDV 62
FT 11
FT CHAIN 1 35
FT SEQUENCE 36 504 AA; 55540 MW; ADAB4B94DD4F02E1 CRC64;
Db 22 CPEFLCLLCC-----GAAGRP-----EYMPSSA-----GSGDRHSAGVOLLH-PW 63
QY 63 FSVHKEKRTLIFR---VRQGG-----QSEGEYERLSTLDRCATLALQVTPQDE 111
Db 64 ECLHLNRVYLCQPRPRNSGEAVRHNASGVRIDETETYSERLSVGEEDA-LSISKVTRQDN 122
QY 112 -RIFLCQ--GKRPRSOEYRIOLRVYKAPPEPNIOVNLGIPVNSKEPEEVAATCGRNGY 168
Db 123 ARTICQVAGDSQGVESRTELTYKIPAPPEITPNSAGIPAGSNDMLKIAQCTSENSFP 162
QY 169 IPQVIWYKNGRPKLEKNRVHIQSSQTVSSGILYTLQSLKADLYEKDKDAQFVCELNVR 228
Db 183 SPNTITWYKNGEPILQEDDKRILTLTVRESNGITVSTLFSKYTRDRNSLPHCTVHW 242
QY 229 LPSNHHKESSEYVVPFYPTPEKWLVEVP-VGMLKEGDRVEIRCLADGNPPHFSISKQ 287
Db 243 LQGMRTKDSFRVAVTFEYPTPEHVELRVATNAGIVKEGDVLCVCDADGNPAPVFSFR 302
QY 288 --NPSTREAEETNDNGVLYLEPARKEHSGRYECQGLDLDTMISLSEPOELLVNVSD 345
Db 303 ELGSMQDMSTSLADTNDGVLMLHNVSSSGLYRCQTLDDMTQHEGD-VELVYNTIEG 361
QY 346 VRV--SPAAPERQSGSLTLTCEAESSQDLEFQWLREBTQVLEGRPVQLDHLKREAG 403
Db 362 VOVMKESPSP-LHGGDSVRLSCTAHSPVKLDYQW-RDARGRKVAEGNQLLTNLTFETSS 419
QY 404 GYRCVAVSPSPICLNKTOLVNVAIFGPPMAFKERKVKENNVNLNLSCEASGHPRTIS 463
Db 420 NESCVRKARSPGLEQSKQVAAVAKGPRIVAISAPLYRQDEVINLTCKAIAFPQSPFH 479
QY 464 MNVNGTASEDQDQRIYSLTNLVLPPELLTETGECTASNDLGKNTSIFLELVNLTTLT 523
Db 480 MSVAGTTHEY-MENQHMASNLTVAVSHDLRAGAMCRVSNALGSE-----KHQLDLQK 533
QY 524 PDSMTTGLSTSTASPTFRANSTSTERKLPEPESRGVYVAVYVCLIVLAVLGAIVFLY 563
Db 534 PS-----ESKGIIIVAIIVCLIVLAVLSIIYELH 563
QY 584 KKGLPCGRSGKQEI 598
Db 564 KKGIISGGRSGKODI 578
RESULT 11
Q98923 PRELIMINARY; PRT; 504 AA.
ID Q98923;
AC Q98923;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H.B190V+; TISSUE=THYMUS;
RA MBDLINE=9713433; PubMed=8978830;
RT Valinlo O., Dunon D., Aissi F., Dandy J.P., McNagay K.M., Imhof B.A.;
RT "HEMCM, an adhesion molecule expressed by c-kit+ progenitors.";
RL J. Cell Biol. 135:1655-1668(1996).
DR EMBL; Y08854; CA70079.1;
DR InterPro; IPR003596; Ig_L2;
DR InterPro; IPR003600; Ig_Like;
DR InterPro; IPR003006; Ig_MHC;
DR Pfam; PF00047; Ig_5;
DR SMART; SM00408; Igc2; 1;
DR SMART; SM00410; Ig_Like; 2;
KW Immunoglobulin domain; Signal.
```

```
FT SIGNAL 1 35
FT CHAIN 36 504
FT SEQUENCE 504 AA; 55540 MW; ADAB4B94DD4F02E1 CRC64;
POTENTIAL.
POTENTIAL.
```

Query Match 22.2%; Score 747; DB 13; Length 504;
Best Local Similarity 35.9%; Pred. No. 7,1e-53;
Matches 172; Conservative 91; Mismatches 184; Indels 32; Gaps 13;

```
QY 8 CAFLAACCCPRVAGVGEAEOPAPELVEVNGSTALLKCGLS-QSQGNLSHVDVSHV 66
Db 22 CPEFLCLLCC-----GAAGRLVYMPAYLEVEIGSTARLESFSPGNASFTSIEMFYVN 76
QY 67 KEKRTLIFVRQGG-----QSEPEYERLSTLDRCATLALQVTPQDE-RIFLCQ 117
Db 77 RRP-----LRGEAVRHNASGVRIDETETYSERLSVGEEDA-LSISKVTRQDNARTFICQ 129
QY 118 --GKRPRSOEYRIOLRVYKAPPEPNIOVNLGIPVNSKEPEEVAATCGRNGYPIQVITW 175
Db 130 VGADSQGVESRTELTYKIPAPPEITPNSAGIPAGSNDMLKIAQCTSENSFPSPNITW 189
QY 176 KNGRPLKEKNRVHIQSSQTVSSGILYTLQSLKADLYEKDKDAQFVCELNVRPSGNHM 235
Db 190 KNGEPLQEDDKRILTLTVRESNGITVSTLFSKYTRDRNSLPHCTVHWLQGMRT 249
QY 236 KESREYVVPFYPTPEKWLVEVP-VGMLKEGDRVEIRCLADGNPPHFSISKQ--NPSTR 292
Db 250 KDSPRNVTVPFTEHVELRVATNAGIVKEGDVLCVCDADGNPAPVFSFRRELQDSMQ 309
QY 293 EAEETTNDNGVLYLEPARKEHSGRYECQGLDLDTMISLSEPOELLVNVSDVRV--SP 350
Db 310 DMTSLADTNDGVLMLHNVSSSGLYRCQTLDDMTQHEGD-VELVYNTIEGVQVKMEP 368
QY 351 AAPEREGSILTLTCEAESSQDLEFQWLREBTQVLEGRPVQLDHLKREAGGRCVYAS 410
Db 369 SSP-LHGGDSVRLSCTAHSPVKLDYQW-RDARGRVAGSNQLLTNLTFETSSNFSRVK 426
QY 411 VPSIPGLNTQLVNVAIFGPPMAFKERKVKENNVNLNLSCEASGHPRTISNVNCT 469
Db 427 ARSVPLEQSKQVAAVAKGPRIVAISAPLYRQDEVINLTCKAIAFPQSPFHSINGT 485
```

```
RESULT 12
Q9MZ08 PRELIMINARY; PRT; 628 AA.
ID Q9MZ08;
AC Q9MZ08;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.
OC NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20363335; PubMed=10908043;
RA Boado R.J., Li J.Y., Pardridge W.M.;
RT "Selective Lutheran glycoprotein gene expression at the blood-brain
RT barrier in normal brain and in human brain tumors.";
RL J. Cereb. Blood Flow Metab. 20:1096-1102(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Shusta E.V., Boado R.J., Pardridge W.M.;
RT "Vascular Proteomics and Subtractive Antibody Expression Cloning.";
RL Mol. Cell. Proteomics 0:0-0(2001).
DR EMBL; AF270512; AAF81749.2;
SQ SEQUENCE 628 AA; 68002 MW; 2128B1F1B73E6A93 CRC64;
```

Query Match 19.2%; Score 645.5; DB 6; Length 628;

Best Local Similarity 30.3%; Pred. No. 2.2e-44;
Matches 205; Conservative 106; Mismatches 263; Indels 103; Gaps 28

OY		PRVLCAFILIAACCCCPRVAGVPGEAOPAPBELVEVVGSTALCKGLSCGSLNSHVD	62
Db	14	PRLLVALILLAA-----PEGRKAEVLSPPLVEVWRGGSVTLDCS-PLGTIDHYEMLEW	66
OY	63	FSVHK-----EKRLLIFVRQ--GGQGSEPGEGYECGRLSLDRGATLALTGYTPDD	110
Db	67	FLVRSRGAHRRLASAEIIRGSELIRDKELNKSRRSP----YQLDSQGR---LVLPKAOVG	119
OY	111	ERIEICCGK--RPSSQRYRIOLRYKAKPEEPNIQVNLGIPIVMSKBEEEVATCVGNYP	168
Db	120	ERDVCYVVKAGACTAETATKLAKFAKPEAPEVSPNKGLISVMDRFQAQETATCSSRGNP	179
OY	169	IPOVIWKNGRPLEKE-EKNRVHIQSSQT-ESSGLTTLQSIIILKAOLYRKDKDAQCYCEL	225
Db	180	APOIMMYRNQCPLAVPLEVNSEGVMTRTVREABAGLILSLITLYLRHKPDREASFHCIV	239
OY	226	NIRLPSSGHMK-ESERTVYPPFYTEKY-WLEVFP--GMKGEDRYELRCIADGNPP	280
Db	240	HYYLPAGOHGLDPSFESLTLYHTPEHVLFMLGSGSTAEGWVRBGDSVOLTCOGDSSPTP	299
OY	281	HFSI-----SKONPSTRAEETTDNDGVILEPARKESHGREYCQGLDLDTMISTL-TSEP	335
Db	300	EYTFPMIADKREDYLK-----TSLEGNTLIERQRNQSGTYGCGRVEDFDPVEDALSKT	353
OY	336	QELLNVYSDVVRSPAAPEROE--GSSLTITCAESSQDLEROMLREETGOYLERGPVL	392
Db	354	LELRVAAYLSDLELS-AGEBLSLPLHNSTVTYCARLPTPLYMTWDSA--FMGEDPTL	409
OY	393	QLHDIKFRAGGRCVAVSPISPLANTOLVNVAIFPPMPMAFEREK-----FWKENMV	447
Db	410	SLHSVTEDSACTYCEAMRPIRLPLSTRSRLLYQGTPEPKAKENQPKRAGSWTEDEV	469
OY	448	LNLSCBSGAPRPITSMN-VNGTASEDDQBP-RVLSLAVLYTPELETGVECTASNDL	505
Db	470	-TLICYARGYKPKLITVSOLGSPTEPAFGQGVSSSLTKIKTYSALSODGVCEASNPL	528
OY	506	GKNTSILFELVNLTTLPDSNTTGTLSTASPHTANSTSTPERKLPEDESNGVYAV	565
Db	529	G-----NHIVHFEGTVAPOT-----SQAGVAMV	554
OY	566	IVCILVLAIVGLVLYFLYKKGLPCRRSGKOETILPPRSKSELVEYKSDKLEPEGGLLO	625
Db	555	AISVALLLVAVAFYCMKRKRKRPCQQMG--EKSSPPGPCPL--SHSGSQRPREGTILM	610
OY	626	GSS--GDKRAPCDGOK	640
Db	611	GSASGAGKHGSGGFGE	627
RESULT	13		
O9UKB2	O9UKB2	PRELIMINARY;	PRT; 622 AA.
AC	O9UKB2:		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, last annotation update)		
DE	LUHERAN GLYCOPROTEIN.		
GN	LU.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxId=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=129/SV;		
RA	Lee G., Willig T.-N., Parsons S.F., Anstee D.J., Mohandas N.,		
RA	Chasis J.A.;		
RT	"Mouse Lutheran Glycoprotein Gene."		
RL	submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.		
EMBL	AF246667; AF61742.1; -		

DR MGD: MGI:1929940; Lu.
DR Interpro: IPR003598; Ig_C2.
DR Interpro: IPR003600; Ig_like.
DR Interpro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig_5.
DR SMART: SM00406; IGC2; 2.
DR SMART: SM00410; Ig_like; 2.
DR Immunoglobulin domain.
KW SEQUENCE 622 AA; 67544 MW; 3A877738D96f66CD9 CRC64;

Query Match	18.4%	Score 618	DB 11	Length 622
Best Local Similarity	28.5%	Pred. No. 4e-42		
Matches 191	Conservative 102	Mismatches 275	Indels 102	Gaps 25

QY	10	FLLAACCCCPVAGVPGABEAPAELEYEVENGTALLKGLSOSQGNLH--VDHF-----	63
Db	15	FLLSG-----YSGAQLAHVSVPPREYVMKGEQVALDC--TPREPHETYLEWLVDS	65
QY	64	-----SVHKKATLLEFRVROGQGSSEPEVEQRLSLDORATATLALQVTPQDERPL	115
Db	66	TGARHRLASVEPQGSSEFLGTV-HSLGRVPREYDNR-----GLVIAVAGQGDGRDY	117
QY	116	COGK--RRRSOEYRLQLRVYAPAREPNIQVNPGLGIPVNSKEPEEVATCVGRNGYPIRQY	173
Db	118	CYVKKGAAGTSEANSSVVFATPBDTEVSPKGTLSVMDQFADELACSSNNKNPVRILT	177
QY	174	WYKNRPLK-----EENKRVHISQSTVESSGITLQSLIKALQVKEKDQAFQCELYVRP	230
Db	178	WYRNQGRLEVEVMENQKQKITIRIVREASGILSTLSTLYLRHKDDRDADFHCAHYDLR	237
QY	231	SGNHMK-ESREYTVVVFPTERY--WLEVEV--GALKEDRVETIKLADGNPRPHFSIS	285
Db	238	SCQHRDLDSHFFRLTLAHPTHEVFEWVGSPSTTEGWREGDVAQQLCQGGSGSPSEYF	297
QY	286	KQNPSTREAEETTYNDNGVULVLEPARKHSGRYEC--QGDLDTMISLSEPOELLVNTY	343
Db	296	RQO-GTQE-EQUNVLKKNLLEGVNRNQGITYGCRVEDYDADBEQVLAK-LKLHAYTL	354
QY	344	SVYRVSAPAR-----QGGSSILTLCEAFSSQDDEFQWLAEEFGQVLERPVLQMLDKR	399
Db	355	DFLEIS-APELQVLELNSSTTVNCSKRGILPTVYKMTDSV--TLADSPMLTDSVTF	410
QY	400	EAGGGYRCVASYPSIPGLNRLQLVNVAIFGPWM-----AFKERYWVKENMYLNSCEA	454
Db	411	DSAGITYTEASPTPYPLLSRTQSPQLLYQGAPELKPMEIMPKSGNSMTBEDEVN-LTCSA	469
QY	455	SGHPPTISMVNWNGTASQDDDPOR--VLSTLVNLYTPLELLEFGVECTASNDYCKNTSIL	512
Db	470	RQFPKPKLTWSRGQTPRAEPPEFGKMKSSIAMKYVTSALSREVSCSEANIHGKKGHNF	529
QY	513	FLELVNLTLLPDSNTTGLSTSPASPHTRANSSTERKLPPEPSRCGVTVVAVIYCLVY	572
Db	530	HGGSV-----APQNAQGVANMAVAASVGLL	555
QY	573	AVLGAVALYFLYKKGKLPGRSGKOEITLPSPRSKSELVEVYKSKLPEPMGLQD--SSGD	630
Db	556	LLVVAFAFCMRKGRHPPCCR--RAEKGAPARBEPL--SHSGSRPEPTGLIMGSPSGG	611
QY	631	KRAPDQGEK 640	
Db	612	RGGSGGFDE 621	
RESULT	14		
Q9R069			
ID	Q9R069	PRELIMINARY;	PRT: 622 AA.
AC	Q9R069:		
DT	01-MAY-2000 (Tremblrel_13, Created)		
DT	01-MAY-2000 (Tremblrel_13, Last annotation update)		
DT	01-DEC-2001 (Tremblrel_19, Last annotation update)		
DE	LUTHERAN GLYCOPROTEIN.		
GN	GPU OR LU.		

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Rahnel C., Colin Y., Goossens D., Gane P., El Nemer W., Carton J.P.,
RA Le Van Kim C.;
RT "Characterization of a mouse laminin receptor gene homologous of the
RT human blood group Lutheran gene";
RL Immunogenetics 0:0-0(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Parsons S.F., Lee G., Chasis J.A., Tanner M.J.A., Anstee D.J.;
RT "Mouse Lutheran glycoprotein";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA MEDLINE=21398054; PubMed=11507772;
RA Moulson C.L., Li C., Miner J.H.;
RT "Localization of Lutheran, a novel laminin receptor, in normal,
RT knockout, and transgenic mice suggests an interaction with laminin
RT alpha5 in vivo";
RL Dev. Dyn. 222:101-114(2001).
DR EMBL: AF109160; AAF14226.1;
DR EMBL: AF221507; AAF34657.1;
DR EMBL: AF346663; AAK83237.1;
DR MGI: 1929940; Lu.
DR InterPro: IPR003598; Ig_C2.
DR InterPro: IPR003600; Ig_Like.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig; 5.
DR SMART: SM00408; IgC2; 2.
DR SMART: SM00410; Ig_Like; 2.
KM Immunoglobulin domain.
SQ SEQUENCE 622 AA; 67670 MW; 257903F08D47EB4A CRC64;

Query Match 18.3%; Score 614; DB 11; Length 622;
Best Local Similarity 28.4%; Pred. No. 8.5e-42;
Matches 190; Conservative 102; Mismatches 276; Indels 102; Gaps 25;

QY 10 FILAACCCPRVAGVPEAEOPAPELVEVEGSTALICKGLSOGNLSH--VDWF---- 63
DB 15 FILSG-----YSGAQAELHVSVPPEVEMRGEQVALDC--TPREHEHYVLEWFLVDG 65
QY 64 -----SVHKEKRTLIFRVNQGOGSEPEYERLSLDRGATLTALQVTPDDEFL 115
DB 66 TGAHRRLASVPEOGSEFLGTV-HSLGRVPEYVDSR-----GRLYIAKVQVGDGRDY 117
QY 116 CQCK--RPRSOERYIQLRVYKAPPEPNIOVPLGIPVNSKEPEEVAFCVGRNGYPIPOVI 173
DB 118 CVYKAGAGAGTSATSSVAFATPEDETVSPNKGTLISVMDQAEIATCSSNNGNPVRIT 177
QY 174 WYKNGRPLK---EKKRNVHIQSSQTVESGLTYLTOSTLKAQLVKEDKDAQFYCELANRPL 230
DB 178 WYKNGQRLVEPMEVNOGYITIRTVREASGLYSLTLYRLKHDRDANFHCACAHYDLP 237
QY 231 SGNHMK-ESRETVVVFVPTREKV--WLEVEPV--GMKKEGRVRIICLADGNPPPHSIS 285
DB 238 SGOGRILDSHTFRLTLHYPTHEHFVWGSPTTEGWNREGDAVQLCGDGSPPSPYSF 297
QY 286 KQNSTREAEETTNDNGVLVLEPARKHSGRYEC--QGIDLDTMISLSEPOELLVNVY 343
DB 298 RQO-GTQE-DQLNVNMLKGNLTTLERVHNOGIYGCREDYDADAEVOLVAK-LKLHVAYL 354
QY 344 SDVAVSPAAPER---QEGSSLTTCGAEASSDLEFQMLKEEIGVLENGPVYQLDILK 399
DB 355 DPELIS--VEELFVFLNSSLSTVAVNCARGLPTPTVAKWKDSV--TLAOGPMLISQSVTF 410
QY 400 EAGGRCVAVSVPISIPGLNLTQVNALEFGPPVM---AFKERRKYVKEENMLINISCEA 454
DB 411 DSAGTVCYCASTPTVPLRLSTQSGFOLYQCAPLKKPRNEIMPKSGNSMTBEDVYM-LTCSA 469

QY 455 SGHPRPTISMNVNGTASBEDDQBPQ--VLSTLNVLTPELLETGYECTASNDLCKNTSIL 512
DB 470 RGFPEPKLTWSQGDPPAPPEPGEGRGWKSSLMVKTALSREGVSCASNHGKRGV 529
QY 513 FLELVNLTLPDPSNTTGLSTASPHTRANSTERTKLPEPESRGVYVAVICIVL 572
DB 530 HFGSV-----APQTAQGVAVNAVAVSGLL 555
QY 573 AVIGAVLYFLYKKGKLPCKRRSGKEITLTPSRKSELVEYKSDKLPEEMGLQG--SSGD 630
DB 556 LTVAAFYCMRRKRGRCGR--RAEKGAPPAREPEL--SHSGSERPEHTGILMGSPSGG 611
QY 631 KRAPDQGEK 640
DB 612 RGGSGGFDE 621

RESULT 15
Q9ESS5 PRELIMINARY; PRT; 622 AA.
AC Q9ESS5;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
GN LUTHERAN ANTIGEN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=BAH/C;
RA Taira E., Okumura S., Miki N.;
RT "mouse Lutheran antigen.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB035511; BAB16053.1;
DR InterPro: IPR003599; Ig_C2.
DR InterPro: IPR003598; Ig_C2.
DR InterPro: IPR003600; Ig_Like.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig; 5.
DR SMART: SM00409; Ig; 5.
DR SMART: SM00408; IgC2; 3.
DR SMART: SM00410; Ig_Like; 3.
KM Immunoglobulin domain.
SQ SEQUENCE 622 AA; 67543 MW; 3264B7490E28428 CRC64;

Query Match 18.2%; Score 613; DB 11; Length 622;
Best Local Similarity 28.4%; Pred. No. 1e-41;
Matches 190; Conservative 102; Mismatches 276; Indels 102; Gaps 25;

QY 10 FILAACCCPRVAGVGEAEOPAPELVEVEGSTALICKGLSOGNLSH--VDWF---- 63
DB 15 FILSG-----YSGAQAELHVSVPPEVEMRGEQVALDC--TPREHEHYVLEWFLVDG 65
QY 64 -----SVHKEKRTLIFRVNQGOGSEPEYERLSLDRGATLTALQVTPDDEFL 115
DB 66 TGAHRRLASVPEOGSEFLGTV-HSLGRVPEYVDSR-----GRLYIAKVQVGDGRDY 117
QY 116 CQCK--RPRSOERYIQLRVYKAPPEPNIOVPLGIPVNSKEPEEVAFCVGRNGYPIPOVI 173
DB 118 CVYKAGAGAGTSATSSVAFATPEDETVSPNKGTLISVMDQAEIATCSSNNGNPVRIT 177
QY 174 WYKNGRPLK---EKKRNVHIQSSQTVESGLTYLTOSTLKAQLVKEDKDAQFYCELANRPL 230
DB 178 WYKNGQRLVEPMEVNOGYITIRTVREASGLYSLTLYRLKHDRDANFHCACAHYDLP 237
QY 231 SGNHMK-ESRETVVVFVPTREKV--WLEVEPV--GMKKEGRVRIICLADGNPPPHSIS 285
DB 238 SGOGRILDSHTFRLTLHYPTHEHFVWGSPTTEGWNREGDAVQLCGDGSPPSPYSF 297

```

OY 286 KNPSTRAEETNDNGVILEPARKHSGRYEC--OGLDPMISLSPEOLLVNV 343
    :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 298 RQO-CTQE-EQLNVNKLKGNLTLEGVHRNOSGITGCRVEDYDADEVOQLVK-LKLVAYL 354
OY 344 SDVRVSPAPER---QEGSSLTITCEAESQDLEFQWLRETEGVLERGVYQLHDLKR 399
    :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 355 DPLEIS--VPELEFVFLNSSSTVNCARGLPTPARWTKDSV--TLADGPMLSLSQSYTF 410
OY 400 EAGGGRCAVAVPSIPGLNRTQLVNVAIFGPPWM---AFKERKVVYKENVVNLSCFA 454
    :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 411 DSAGFYTCESASTPVPLSLRTQSFLIVQGAPELKPNEIMPKSGNSWTEGDEVN-LTCSA 469
OY 455 SGHPPTISMNVNCTASBDDDPOR--VLSTLNVLYPPELLETGVECTASNDLGKNTSIL 512
    :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 470 RGFPEPKLWISORGDTPAPAPPEEGRGKSSSLMVKVTSALSREGVSCASNIHGKKGHYF 529
OY 513 FLELVNLTTLTPDSNTTGLSTSTASPHTRANSTSTERKLPEPESRGVIVAVIYCIYL 572
    :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 530 HRGSV-----APQTAQAGVAVMAVAVSVGLL 555
OY 573 AVLGAVALFLYKKGKLPGRSGKQETTLPPSRKSELVVEVKSDDLPEMGILLG--SSGD 630
    :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 556 LLYVAAFYCMRRKGRPGCCR--RAEKGAAPAREPEL--SHSGSERPEHTGLMGSPSGG 611
OY 631 KRAPGDOGEK 640
    :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 612 RCGSGGFGDE 621

```

Search completed: June 28, 2002, 10:50:26
 Job time: 276 sec

1617 CTTCCACCTGCGATCTCTATACAGAGCAGACACACTCCACAGAGAGA 1666
551 LysLeuProGluProGluSerArgGlyAlaValIleValAlaValIleVal 567
1667 AACGTCGCGAGCCGAGAGCGGGGCGGTCTCATCTGCTGTGATGCT 1716
567 LysIleLeuValIleValAlaValIleValAlaValIleValIleVal 584
1717 GTCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1766
584 LysIleLeuValIleValAlaValIleValAlaValIleValAlaVal 600
1767 AGAAGGAGCAAGCTGCGTGCAGAGCGCTCAGAGAGAGAGATCAGCTG 1816
601 ProProSerArgLysSerGluLeuValAlaValLysSerAspLysLe 617
1817 CCCCCTGCTGTAAGAGCAACTGTTGTAAGTAACTAAGTAACTAAGCT 1866
617 UPProGluGluMetGlyLeuLeuGlnGlySerSerGlyAspLysArgAlaP 634
1867 CCCAGAGAGATGGGCTCTCTGAGGCGAGCGGTGACAGAGAGGCTC 1916
634 roGlyAspGlnGlyLysTyrIleAspLeuArgHis 646
1917 CCGGAGACCAAGGAGAAATTCATCTGATGAGGCT 1954
seq_name: qb_pr:AF089868
seq_documentation_block:
LOCUS AF089868 3335 bp mRNA linear PRI 04-MAR-1999
DEFINITION Homo sapiens cell surface glycoprotein PNH2 precursor, mRNA,
complete cds.
ACCESSION AF089868
VERSION AF089868.1 GI:4336423
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
AUTHORS Gul, L., Chang, L., Browne, P. V. and Heibel, R. P.
TITLE PNH2 from human umbilical vein endothelial cells
JOURNAL Unpublished
2 (bases 1 to 3335)
REVIEWER Gul, L., Chang, L., Browne, P. V. and Heibel, R. P.
AUTHORS Direct Submission
TITLE Submitted (02-SEP-1998) Medicine, University of Minnesota, Box 480
JOURNAL UMMC, 420 Delaware Street SE, Minneapolis, MN 55455, USA
FEATURES
source
1..3335
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="umbilical vein endothelial"
/note="Primary culture"
24..1964
/note="PNH2 antigen"
/codon_start=1
/product="cell surface glycoprotein PNH2 precursor"
/protein_id="A01779.1"
/db_xref="GI:4336424"
translation="MGIPRLVCAFLAACCCPRVAGPGEADPABELVEVGSTAL
LKCGLSOGNLSHWDFSVHREKRTLLFRQCGOGESEGEYORLSQDRATLA
LQVTPQDERFLQGRKRPSEYRQLRVYAPPEPNIQNPGLIVNSKEPEVAT
CVGRNGPPIQVLIYKNGRLKEKRVHIOSSQVESGLYTLKQIVLEEDD
AOFCELENRLPSGNHMKESREVTVPYPTKVMLEVPAGMLKEGDRVEICLAG
NPPEFISKONPSTRAEETNDNGVILTPARKBSGRTGECGLDOLTEGSPVL
NPOLLVTVSVRSPAPAREGSSSLITLCASOSBLEOMLREEDDOLTEGSPVL
OLHLKRGAGGACVAVSPSTIGRLROLVLAIEGPMAAFKERYVMKENVLT
SCASGHPRTISMVNGTASPDOPORVLTIVLVYTPLELGGVCAKSNLIGN
TSILFLIVLTLPDSNTGISTSAISPPRANSTSRKLPAPESRGVYVNT
VCLIVLANIVLTFLYKCKLPCRSKSGKQETITPPRSKSELVYVNSDKLPKML
GGSGDKRAFDGEGKTIIDLR"

BASE COUNT 809 a 921 c 941 g 664 t
ORIGIN
alignment_scores:
Quality: 3356.00 Length: 646
Ratio: 5.203 Gaps: 0
Percent Similarity: 99.845 Percent Identity: 99.845
alignment_block:
US-09-653-961-2 x AF089868 ..
Align seq 1/1 to: AF089868 from: 1 to: 3335
1 MetGlyLeuProArgLeuValLysAlaPheLeuLeuAlaLacysCysCys 17
24 ATGGGCTTCCAGAGCTGTGCTGCGCTTCTTGCTGCGCGCTGCTGCTG 73
17 scyProArgValAlaGlyValProGlyGluAlaGluGlnProAlaProG 34
74 CTGTCTCTGCGCTGCGGCTGTGCGCGAGAGCTGAGCAGCTGCGCTG 123
34 ILeuValGlyValGlyValGlySerThrAlaLeuLeuLysCysGlyLeu 50
124 AGCTGTGAGGTGAAGTGGGAGCAGCAGCCTTCTTGAAAGTGGGCTC 173
51 SerGlnSerGlnGlyAsnLeuSerHisValAspTyrPheSerValHisLy 67
174 TCCAGTCCCAAGGCAACCTCAGCCAGCTGACCTGCTTTCTGTCACAA 223
67 scLysArgThrLeuIlePheArgValArgGlnGlnGlnGlnSerG 84
224 GGAGAAGCGGAGCGCTATCTGCTGCGCGCGCGCGCGCGCGAGAGCG 273
84 IUProGlyGluTyrGluGlnArgLeuSerLeuLysPheGlyAlaThr 100
274 AACCTGGGAGTACGAGCAGCGCTGCTGCTGCGCGCGCGCGCGAGAG 323
101 LeuAlaLeuThrGlnValThrProGlnAspGluArgIlePheLeuGly 117
324 CTGGCCCTGACCTCAGTACGACCCCAAGAGCAGCAGCAGCTCTCTG 373
117 nGlyLysArgProArgSerGlnGluTyrArgIleGlnLeuArgValTyr 134
374 GGGCAAGCGCCCTGCTGCGCGAGGTACCGCATCCAGCTCCGCTAC 423
134 yAlaProGluGluProAsnIleGlnValAsnProLeuGlyIleProVal 150
424 AAGCTCCGAGAGGCAACATCCAGGTCAACCCCTGCGCATCTCTG 473
151 AsnSerLysGluProGluGluValAlaThrCysValGlyArgAsnGly 167
474 AACAGTAAAGAGCCTGAGAGAGTGTGCTGCTGCTGCTGCTGCTG 523
167 rProIleProGluValIleTyrTyrLysAsnGlyArgProLeuLysGlu 184
524 CCCATCTCTCAGTACATCTGTGTCAGAAATGGCCGCTCTGAAAGAG 573
184 IuLysAsnArgValHisIleGlnSerSerGlnThrValGluSerGly 200
574 AGAAGAACCGGCTCCACATTCAGTGCACCACTGTGAGAGAGAGT 623
201 LeuTyrThrLeuGlnSerIleLeuLysAlaGlnLeuValLysGluAsp 217
624 TTGTACACCTTCAGAGATTCCTAAGAGCAGCAGCTGTAAAGAGACA 673
217 sAspAlaGlnPheTyrCysGluLeuAsnTyrAlaGluProSerGlyAsn 234
674 AGATGCCAGTTTACTGTGAGCTCACTACCGGCTGCCAGAGGAAAC 723
234 IsetLysGluSerArgIleValThrValProValPheTyrProGlu 250
724 ACATGAAGAGTCCAGGAGATCAACCTCTCTTTTCAACCGAGAGAA 773

251 LysValIrrPLeuGluValAlGluProValAlGlyMetLeuLysGluGluAspAr 267
|||||
774 AAAGTGTGGTGGAG 823
267 GValGluIleArgCysLeuAlaAspGlyAsnProProProlHisPheSerI 284
|||||
824 CGTGGAAATCAGAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 873
284 LserLysGluAsnProSerThrArgGluAlaGluGluGluThrThrAsn 300
|||||
874 TCAGACACAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 923
301 AspAsnGlyValLeuValLeuGluProAlaArgLysGluHisSerGlyAr 317
|||||
924 GACAGCGGGGTCTGCTGCTGAGACCTGCCCGAAGAGAAACAGAGTGGGG 973
317 gTArgLysGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 334
|||||
974 CTATGAAATGTCAAGGCTGGACTGGACCTGGACACCATATGCTGCTGAGTG 1023
334 LuProGluGluLeuLeuValAsnArgValSerAspValArgValSerPro 350
|||||
1024 AACCAAGAGAACTACTGCTGAACTATGCTGACGCTCGAGTGAATGCC 1073
351 AlaAlaProGluArgGluGluGluGluGluGluGluGluGluGluGluGlu 367
|||||
1074 GCAGCCCTGAG 1123
367 agLysSerGluAspLeuGluPheGluIrrPLeuArgGluIrrArgLys 384
|||||
1124 AGAGATAGCCAGAGACTCGAGTTCAGTGGCTGAGAGAGAGAGAGAGAG 1173
384 LuValLeuGluIrrArgLysProValLeuGluIrrLeuHisAspLeuLysArgLys 400
|||||
1174 AGGTGTGGAAG 1223
401 AlaGluGlyGlyTyrArgCysValAlaSerValProSerIleArgGlyLe 417
|||||
1224 GCAG 1273
417 uAsnArgTrrGluLeuValAsnValAlaIlePheGlyProProTrrPheLa 434
|||||
1274 GAACCCACACAGCTGTGTCAACGTCGCAATTTTGGCCCCCTGTGAGATGG 1323
434 LePheLysGluArgLysValIrrValIrrLysGluAsnMetValLeuAsnLeu 450
|||||
1324 CATTCAG 1373
451 SerCysGluAlaSerGlyHisProArgProThrIleSerTrrPAsnValAs 467
|||||
1374 TTTTGTGAAGCGTCAGGGCACCAGCCGCGCCACATCTCTGTGAACGTCAA 1423
467 ngLysTrrAlaSerGluGluGluGluGluGluGluGluGluGluGluGlu 484
|||||
1424 CGCAG 1473
484 euAsnValLeuValThrProGluLeuGluGluGluGluGluGluGluGlu 500
|||||
1474 TGAATCTCTGTCAGCCCGAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAG 1523
501 AlaSerAsnAspLeuGlyLysAsnTrrSerIleLeuPheLeuGluLeuVal 517
|||||
1524 GCCTCCAAAG 1573
517 LuAsnLeuThrThrLeuThrProAspSerAsnTrrThrThrArgLysLeuSer 534
|||||
1574 CAATTTAAACCACTCTCAACAGAGCTCAACAGAGAGAGAGAGAGAGAGAGAG 1623
534 hTrSerThrAlaSerProHisThrArgValAsnSerThrSerThrGluArg 550
|||||
1624 CTTCACCTCCAGTCTCTATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1673

551 LysLeuProGluIrrProGluSerArgGlyValValIleValAlaValIleVal 567
|||||
1674 AAAGTGTGGTGGAG 1723
567 LysIleLeuValLeuAlaValLeuGlyAlaValLeuTyrPheLeuTyrL 584
|||||
1724 GTGCATCTGTGGTCTGGGGGTGGGGGTGGGGGTGGGGGTGGGGGTGGGGGT 1773
584 LysLysGlyLysLeuProCysArgArgSerGlyLysGluGluIrrLeuLeu 600
|||||
1774 AGAAGGCAAG 1823
601 ProProSerArgLysSerGluLeuValValAlaGluValLysSerAspLysLe 617
|||||
1824 CCCCCTGCTGTGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1873
617 uProGluGluMetGlyLeuLeuGluGluGluGluGluGluGluGluGluGlu 634
|||||
1874 CCCAGAGAGAGATGGGCTCTCTGACAGGAGAGAGAGAGAGAGAGAGAGAG 1923
634 roGlyAspGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 646
|||||
1924 CGGAG 1961
seq_name: gb_pat:AX331551
seq_documentation_block:
LOCUS AX331551 2943 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 2060 from Patent WO0194629.
ACCESSION AX331551
VERSION AX331551.1 GI:18122185
KEYWORDS
SOURCE
ORGANISM
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (sites)
AUTHORS Young,P.B., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
Hortigan,S., Soppet,D.R. and Weiser,Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A, 2060 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
source
1..2943
location/Qualifiers
BASE COUNT 668 a 863 c 858 g 554 t
ORIGIN
alignment_scores:
Quality: 3327.00 Length: 646
Ratio: 5.166 Gaps: 0
Percent Similarity: 99.690 Percent Identity: 98.916
alignment_block:
US-09-653-961-2 x AX331551 ..
Align seg 1/1 to: AX331551 from: 1 to: 2943
1 MetGlyLeuProArgLeuValCysAlaPheLeuLeuAlaIleCysCysCys 17
|||||
8 ATGGGGCTTCCTCCAGAGCTGTGGCTGTGCTTCTGCTGCGCGCTGTGCTG 57
17 SCysProArgValAlaGlyValAlaProGlyGluAlaGluGluGluGluGlu 34
|||||
58 CTGCTCGCGGTGCGGGGTGGGGGTGGGGGTGGGGGTGGGGGTGGGGGTGGGG 107
34 LuLeuValGluValGluValGlySerThrAlaLeuLeuLysGlyGlyLeu 50
|||||
108 AGCTGTGAGAGTGAAGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 157

seq_documentation_block:
LOCUS AX332735 2943 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 3244 from Patent WO0194629.
ACCESSION AX332735
VERSION AX332735.1 GI:18123369
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (sites)
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horrihan, S., Soppet, D. R. and Weaver, Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 3244 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
source location/Qualifiers
1..2943
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 668 a 863 c 858 g 554 t
ORIGIN
alignment_scores:
Quality: 3327.00 Length: 646
Ratio: 5.165 Gaps: 0
Percent Similarity: 99.690 Percent Identity: 98.916
alignment_block:
US-09-653-961-2 x AX332735 ..
Align seg 1/1 to: AX332735 from: 1 to: 2943
1 MetGlyLeuProArgLeuValCysAlaPheLeuLeuAlaAlaCysCysCY 17
|||||
8 ATGGGGCTCCAGAGCTGGTCTGCCCTTCTGGTGGCCGCTGCTGCTG 57
|||||
17 sCyProArgValAlaGlyValProGlyAlaGlyLeuProAlaProG 34
|||||
58 CTGTCTCTCCGTCGCGGGTGTCCCGGAGAGCGTCAGACGCTGGCCTG 107
|||||
34 LuLeuValGluValGluValGlySerThrAlaLeuLeuLysCysGlyLeu 50
|||||
108 ACCTGTGGAGGTGGAAGTGGGCACACAGCCCTCTGAAGTGGCGCTC 157
|||||
51 SerGlnSerGlnGlnAsnLeuSerHisValAspTyrPheSerValHisLy 67
|||||
158 TCCCGAGTCCCAAGCAACTCAGCCATGTGAGCTGGTTTCTGTCCAA 207
|||||
67 sGlyLysArgThrLeuIlePheArgValArgGlnGlyGlnGlnSerG 84
|||||
208 GGAGAGCGGAGCGCTCATCTTCCGTGTGCGCAGGGCCAGGGCCAGAG 257
|||||
84 LuProGlyGlyTyrGlnGlnArgLeuSerLeuGlnAspArgGlyAlaThr 100
|||||
258 AACCTGGGGAGTACGAGCGCGCTCAGCTCCAGAGACAGAGGGCTACT 307
|||||
101 LeuAlaLeuThrGlnValThrProGlnAspGluArgIlePheLeuCysG 117
|||||
308 CTGGGCGCTGACTCAAGTCAACCCCAAGACAGAGCGCATCTTTTGTGCA 357
|||||
117 nGlyLysArgProArgSerGlnGlyTyrArgGlnGlnLeuArgValTyrL 134
|||||
358 GGGCAGACGGCCCTCGGTCCAGAGATACCGCATCCAGCTCCGCTTACA 407
|||||
134 ySalAProGlyLeuProAlaIleGlnValAsnProLeuGlyIleProVal 150
|||||
408 AAGCTCCGAGAGACCACATCCAGGTCAACCCCGGGGCAATCCCTGTG 457
|||||
151 AsnSerLysGluProGlyLeuValAlaThrCysValGlyArgAsnGlyT 167
|||||

458 AACAGTAAGACCTGAGAGAGTCCCTACCTGTGTAGAGAGAAACGGTA 507
167 rProIleProGlnValIleTyrTyrLysAsnGlyArgProLeuLysGluG 184
|||||
508 CCCATTCTCTCAAGTCATCTGTAGACAAAGATGGCGGCTCTGAAGAG 557
|||||
184 LuLysAsnArgValHisIleGlnSerSerGlnThrValGlnSerSerGly 200
|||||
558 AGAAGAACCGGGTCCACTTCATCTGCTCCAGACGTGGAGTGAATGAT 607
|||||
201 LeuTyrThrLeuGlnSerIleLeuLysAlaGlnLeuValLysLysPly 217
|||||
608 TTGTACACCTTGCAGAGATTTCTGAAGGACGAGCTGTAAAGAACACA 657
|||||
217 sAspAlaGlnPheTyrCysGluLeuAspTyrArgLeuProSerGlyAsnH 234
|||||
658 AGATGCCAGTTTACTGTAGCTCAACTACCGGCTGCCAGTGGAGACC 707
|||||
234 lsmetLysGlnSerArgGluValThrValProValPheTyrProThrGlu 250
|||||
708 ACATGAAGAGATCCAGGAAATCACCGTCCGTTTCTACCCGACAGA 757
|||||
251 LysValThrLeuGluValGluProValGlyMetLeuLysGlnGlyAspAr 267
|||||
758 AAGTGTGGCTGGAAAGTGCAGCCCGTGGGAATGCTGAAGGAAGGAGACG 807
|||||
267 gValGluIleArgCysLeuAlaAspGlyAsnProProProHisPheSerI 284
|||||
808 CGTGGAATCAGAGTGTGTGGCTGATGGACACCTCCACACTTTCAGCA 857
|||||
284 lSerLysGlnAsnProSerThrArgGluAlaGlyGluGlnGluThrThrAsn 300
|||||
858 TCAGCAGAGACACCCAGCAGCAGGAGAGAGAGAGAGAACACACCAAC 907
|||||
301 AspAsnGlyValLeuValLeuGluProAlaAspGlyLysSerGlyAr 317
|||||
908 GACAAAGGGGTCTGTGTGTGGAGCCTGCCGGAAGAAACACAGTGGGG 957
|||||
317 gTyrGlnCysGlnLeuAsnLeuAspPheMetLysSerLeuLeuSerG 334
|||||
958 CTATGAATGTGAGCGCTGAACTGTGACACATGATATGCTCTGATAG 1007
|||||
334 LuProGlnGluLeuLeuValAsnTyrValSerAspValArgValSerPro 350
|||||
1008 AACACAGAGAACTACTGTGATATGTCTGACGTCCAGAGAGTCCC 1057
|||||
351 AlaAlaProGluArgGlnGlnGlySerSerLeuThrLeuThrCysGluAl 367
|||||
1058 GCAGCCCTGAGAGACAGAGAGGACAGACCTCACTGACCTGTAGAGC 1107
|||||
367 agLysSerSerGlnAspLeuGluPheGlnThrLeuArgGluGlnThrGlyG 384
|||||
1108 AGAGATAGCCAGGACCTCGAGTTCCAGTGGCTGAGAGAAAGACAGAC 1157
|||||
384 LuValLeuGluArgGlyProValLeuGlnLeuHisAspLeuLysArgGlu 400
|||||
1158 AGGTGTGTGAAGAGGGGCTGTGTCTTCACTGATCACTGAAAGAGGGAG 1207
|||||
401 AlaGlyGlyGlyTyrArgCysValAlaSerValProSerIleProGlyLe 417
|||||
1208 GCAGAGGGGAGTATCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1257
|||||
417 uAsnArgThrGlnLeuValAsnValAlaIlePheGlyProProThrPheVa 434
|||||
1258 GAAACGACACAGAGCTGAGTCAAGCTGGCAATTTTGGCCCGCCCTGGATGG 1307
|||||
434 lAPheLysGluArgLysValTyrPValLysGluAsnMetValLeuAsnLeu 450
|||||
1308 CATTCAGAGAGAGAGAGTGTGGTGAAGAAATAGGTTGATTCG 1357
|||||
451 SerCysGluAlaSerGlyHisProArgProThrIleSerTyrAsnValAs 467
|||||
1358 TCTTGTGAAGCTCAGGGCACCCCGGCCACATCTCCGTGAAGCTCAA 1407
|||||

467 nglyThrAlaSerGluGlnAspGlnAspProGlnAlaValLeuSerThrL 484
 1408 CGGACGGCGAAGTGAAGACAGACAGATCCACAGAGCTCTTCCAGACCC 1457
 484 euAsnValLeuValThrProGluLeuGluGluThrGlyValGlyCysThr 500
 1458 TGAATGTCCTCGTGAACCCGAGCTGTGGAGACAGGTGTGATCAGCAG 1507
 501 AlaSerAsnAspLeuGlyLysAsnThrSerLileuPheLeuGluLeuVal 517
 1508 GCCCTCCAGACCTCGGGCAAAACACAGCATCTCTTCCGTGAGCTGCT 1557
 517 LAsnLeuThrThrLeuThrProAspSerAsnThrThrThrGlyLeuSerT 534
 1558 CAATTTTAAACCCCTCACACAGACTCCACACACACACAGCTGCGCTGCA 1607
 534 hrSerThrAlaSerProHisThrArgAlaAsnSerThrSerThrGluArg 550
 1608 CTTCACCTGCGAGTCTCATACAGACAGACACACACCTCCACAGAGAGA 1657
 551 LysLeuProGluProGluSerArgGlyValValAlaValAlaValLeuVal 567
 1658 AAGTGGCCGAGCGGAGAGCCGGGGCGGTGTGATCTGCTGATGCT 1707
 567 LysAlaLeuValLeuValLeuValLeuValValLeuValPheLeuVal 584
 1708 GTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1757
 584 YLysGlyLysLeuProGlyArgArgArgArgArgArgArgArgArgArg 600
 1758 AGAAGGCGACAGCTCCCTCAGAGCTCAGAGAGAGAGAGAGAGAGAGAG 1807
 601 ProProSerArgLysSerGluLeuValAlaGlyValLysSerAspLysLe 617
 1808 CCCCCTGCTCTGAGACGACACTGTACTGATGATGATGATGATGATGATG 1857
 617 UPProGluGluMetGlyLeuLeuGlyLysSerGlyAspLysArgAlaP 634
 1858 CCGAAGAGATGGGCTCTCTGAGGAGAGAGAGAGAGAGAGAGAGAGAG 1907
 634 TGGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 646
 1908 CGGAG 1945
 seq_name: gb_pr: H00MUC18A
 seq_documentation_block:
 LOCUS H00MUC18A 2943 bp mRNA linear PRI 16-AUG-1994
 DEFINITION Human isolate Juso MUC18 glycoprotein mRNA (3' variant), complete
 cds.
 ACCESSION M29277
 VERSION M29277.1 GI:530047
 KEYWORDS MUC18 glycoprotein; cell adhesion molecule; immunoglobulin-like
 sequence; integral membrane glycoprotein.
 SOURCE Human CDNA to mRNA.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 2943)
 AUTHORS Lehmann,J.M., Riettmüller,G. and Johnson,J.P.
 TITLE MUC18, a marker of tumor progression in human melanoma, shows
 sequence similarity to the neural cell adhesion molecules of the
 immunoglobulin superfamily
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 86 (24), 9891-9895 (1989)
 MEDLINE 90099368
 REFERENCE 2 (bases 1 to 2943)
 AUTHORS Serz,C., Kirsch,K., Rothbacher,U., Riettmüller,G. and Johnson,J.P.
 TITLE Genomic organization of the melanoma-associated glycoprotein MUC18:
 implications for the evolution of the immunoglobulin domains
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 90 (18), 8514-8518 (1993)
 MEDLINE 93391384
 REFERENCE 3 (bases 1 to 2943)

AUTHORS Johnson,J.P.
 TITLE Direct Submission
 JOURNAL Submitted (06-OCT-1989) Judith P. Johnson, Institute for
 Immunology, University of Munich, Goethestrasse 31, Munich, Germany
 80336
 COMMENT On Aug 16, 1994 this sequence version replaced gi:188859.
 FEATURES
 source
 location/Qualifiers
 1. 2943
 /organism="Homo sapiens"
 /isolate="Juso"
 /db_xref="taxon:9606"
 /clone="drop4.7"
 /cell_line="Mel Juso"
 /tissue_type="melanoma"
 8. 1948
 /codon_start=1
 /product="MUC18 glycoprotein"
 /protein_id="AAA20824.1"
 /db_xref="GI:530048"
 /translation="MGPRIVCAFLAAACCCPRVAVPGAEADPAPELVEVGGSTA
 LKCGLSOSQGNLSHVDWFSYHKERKRLIPRVROGGQSEPEGEORLSLDGATLA
 LTVTPQDRIETLCGRPSOEYRIOLRYKAPEDNIOVPLIPVNSKEPEEYAT
 CVGRNGYPIQYIYKNGRPLEKKNVHIOSSQVSEGLTLOSILKAQVLEKD
 AOPYCELANRILPSGNHMKESRETVPVPEYKRWLEVPVGMLEKGRVIRCLADG
 NPPHFSISKNQPSHREKRETTNDNGVLLEPARHSHGRVHCOAMLDPMISLSE
 POELLVNYSQVRSAPAREDESSLILCEASDSDLEPOMLEETDLENGPVL
 OIHDRKPRGGYRCVAVSPISGLINRLOVLAIFGPPMAFKERKVVENNVGLN
 SCRASGHPPTISMVNGTASDODPORPLSLVNLVLPETLGVCTSNIDGN
 TSLILELVNLTPLIPDSNTTGLSTGKSPRNSKQEIILPSPRSTELVVEKSKLPEEMGL
 VCIIVAVIGAVLYFLYKKKGLPCRRSGKQEIILPSPRSTELVVEKSKLPEEMGL
 OSSGKRRAPGQGEKTYIDLKH"
 8. 91
 92. 1945
 /product="MUC18 glycoprotein"
 122. 373
 /note="immunoglobulin-like (V set) domain I"
 misc.feature
 467. 694
 /note="immunoglobulin-like (V set) domain II"
 misc.feature
 795. 988
 /note="immunoglobulin-like (C2 set) domain"
 misc.feature
 1076. 1243
 /note="immunoglobulin-like (C2 set) domain"
 misc.feature
 1337. 1525
 /note="immunoglobulin-like (C2 set) domain"
 misc.feature
 1337. 1525
 /note="immunoglobulin-like (C2 set) domain"
 BASE COUNT 668 a 863 c 534 t
 ORIGIN
 alignment_scores:
 Quality: 3327.00 Length: 646
 Ratio: 5.166 Gaps: 0
 Percent Similarity: 99.690 Percent Identity: 98.916
 alignment_block:
 us-09-653-961-2 x H00MUC18A ..
 Align seq 1/1 to: H00MUC18A from: 1 to: 2943
 1 MetGlyLeuProArgLeuValCysAlaPheLeuLeuAlaCysCysCys 17
 8 ATGGGGCTTCCAGAGGTGGTGTGCGCTCTCTGCGCCGCTGCTGCTG 57
 17 SCSPProArgValAlaGlyValProGlyGluAlaGluGlnProAlaProG 34
 58 CTGTCTCGGTCGCGCGGCGGTGCCGAGAGCTGAGACAGCTCGGCTG 107
 34 LuleuValGluValGluValGlySerThrAlaLeuLeuLeuLysCysGlyLeu 50
 108 AGCTGGTGGAGGTGGAGGTGGAGACACAGCCCTTCCGAAAGTGGCGCTC 157
 51 SerLysSerGlnGlyAsnLeuSerHisValAspTyrPheSerValHisIly 67
 158 TCCAGTCCCAAGGAGCACTCAGCATGTGAGCTGTTTCTGTCCACAA 207


```
67  sglIuYsArqThrLeuIlePheArqValAArgInglYglInsgIserG 84
   |||
208  GGAGNAGCGACCTCATCTTCGTGGCCAGGACCGAGGACGAGAC 257
   |||
84  lUProglYolUtyrGlInglAArgLeuSerLeuGlnAspArglyAlaThr 100
   |||
258  AACCTGGGAGTACAGAGAGGGGTCCAGCTCCAGGACAGAGGGGCTACT 307
   |||
101  LeuAlaLeuThrGlnValThrProGlnAspGluAArgIlePheLeuGly 117
   |||
308  CTGGCCCTGACTCAAGTACACCCCAAGAGAGCGCATCTTGTGGCCA 357
   |||
117  nglYlYsArqProArqSerGlnolUtyrAArgIleGlnLeuAArgValTyrL 134
   |||
358  GGGCAAGCGCCCTCGCTCCAGAGATACCGCATCCAGCTCCGCTCTACA 407
   |||
134  ySAAlaProglInglUProasnIleGlnValAsnProLeuGlyIleProVal 150
   |||
408  AACCTCCGAGAGAGCAACATCCAGGTACACCCCTGGGCTCGTGTG 457
   |||
151  AsnSerLysGluProglInglUValAlaThrCysValGlyAArgAsnGlyTyr 167
   |||
458  AACAGTAAAGACCTGAGAGAGGTGCTACCTGTGTAGGAGAGAACGGCTA 507
   |||
167  rProIleProglInValIleTrrPtyrLysAsnGlyAArgProLeuLysGlnG 184
   |||
508  CCCCATTCTCTAAAGTATCATCTGTGTACAAAGATGGCCGGCTCTGAAAGAG 557
   |||
184  lUlyAsnAArgValHisIleGlnSerSerGlnThrValGluSerSerGly 200
   |||
558  AGAAGAACCGGGTCCACATTCAGTCCCTCCACAGCTGTGAGTCCAGTGT 607
   |||
201  LeuTyrThrLeuGlnSerIleLeuLysAlaGlnLeuValLysGlnAspLys 217
   |||
608  TTGTACACCTTGACAGATATCTGAAGCACAGCTGGTTAAAGAGACAA 657
   |||
217  sAspAlaGlnPheTyrCysGlnLeuAsnTyrAArgLeuProSerGlyAsnH 234
   |||
658  AGATGCCAGTTTACTGTGTAGCTCAACTCCAGAGCTGGCCAGTGGAGACC 707
   |||
234  lAsMetLysGluSerArqIuValThrValProValPheTyrProThrGlu 250
   |||
708  ACATGAAGAGATCCAGGAGATCACCGCTCGTTTCTTACCGACAGAA 757
   |||
251  LysValTrrPLeuGlnValGluProValGlyMetLeuLysGlnGlyAspAr 267
   |||
758  AAGTGTGGCTGGAAAGTGAAGCCCGTGGGAATCTGAAAGAAAGGAGACCG 807
   |||
267  gValGlnIleAArgCysLeuAlaAspGlyAsnProProPheIlePheSerI 284
   |||
808  CGTGGAAATCAAGTGTCTTGGCTGATGGCAACCTCCACACACTTCAGCA 857
   |||
284  lAserTyrGlnAsnProSerThrAArgIleGlnGlnGlnGlnGlnGlnGln 300
   |||
858  TCAGCAAGCAGAACCCCAACGACGAGGAGGAGAGAGAAAGACAAACCAAC 907
   |||
301  AspAsnGlyValLeuValLeuGlnProAlaAArgLysGlnHisSerGlyTyr 317
   |||
908  GACAAAGGGGTCTGCTGTCTGGAGCTCCCGGAGAGAACCACTAGTGGCG 957
   |||
317  gTyrGlnCysGlnGlnLeuAspLeuAspThrMetLleSerLeuLeuSerG 334
   |||
958  CTTATGATATCTAGGCTTGAACCTTGACACCATGATATCGTGTGAGTGA 1007
   |||
334  lUProglInglUleuLeuValAsnTyrValSerAspValAArgValSerPro 350
   |||
1008  AACCAAGAGAACTACTGTGTAATGTCTGAGAGTCCGAGTGAAGTCC 1057
   |||
351  AlaAlaProglUuAArgInglUlySerSerLeuThrLeuThrCysGlnAl 367
   |||
1058  GCAGCCCTGAGAGAGAGAAAGCACACCTCCACCTGAGACCTGTGTAGGC 1107
   |||

367  aGluSerSerGlnAspLeuGlnUProGlnTrrPLeuAArgGlnGlnTrrGly 384
   |||
1108  AGAGATAGCCAGSACCTGAGATCCAGTGGCTGAGAGAAAGAGACGAGACC 1157
   |||
384  lValLeuGlnUArgGlyProValLeuGlnLeuHisAspLeuLysAArgL 400
   |||
1158  AGGTCTGGAAGAGGGGCTGTGCTTCAAGTTCAGTACCTGGAACGGAG 1207
   |||
401  AlaGlyGlyGlyTyrAArgCysValAlaSerValProSerIleProGlyLe 417
   |||
1208  GCAGAGGGGGCTATGCTGTGCTGGCTGTGTGCTCCAGCATACCGGCT 1257
   |||
417  uAsnAArgThrGlnLeuValAsnValAlaIlePheGlyProProTrrPMe 434
   |||
1258  GAACCGCACACAGCTGTGTCAAGCTGGCCATTTTGGCCCTTGGAGTGG 1307
   |||
434  lA PheLysGlnUArgLysValTrrPValLysGlnAsnMetValLeuAsnLeu 450
   |||
1308  CATTCAGAGAGAGAGAGGTGTGTGGAAGAGATATGCTGTGAATCTG 1357
   |||
451  SerCysGlnUAlaSerGlyHisProArqProThrIleSerTrrPAsnValAs 467
   |||
1358  TCTTGTGAAGGCTCAGGGCACCCCGGCCACATCTCTGTGAGAGTCAA 1407
   |||
467  nglYThrAlaSerGlnGlnAspGlnAspProGlnAArgValLeuSerThrL 484
   |||
1408  CGGACGGGCAAGTGAACAAAGAACCAAGATCCACAGCGAGTCTGAGCACCC 1457
   |||
484  euAsnValLeuValThrProGlnLeuLeuGlnUThrGlyValGluCysThr 500
   |||
1458  TGAATGTCTGTGACCCCGGAGCTGTGGAGACAGGTGTGAATGACAG 1507
   |||
501  AlaSerAsnAspLeuGlnGlyLysAsnThrSerIleLeuPheLeuGlnUleuVal 517
   |||
1508  GCCTCAACAGCACCTGGGCAAAAACACAGCATCTCTCTGTGAGAGTGT 1557
   |||
517  lAsnLeuThrThrLeuThrProAspSerAsnThrThrThrGlyLeuSerT 534
   |||
1558  CAATTATACACCCCTCACACAGACTCCACACACACACAGCGCTCAGCA 1607
   |||
534  hrSerThrAlaSerProHisThrAArgAlaAsnSerThrSerThrGlnUArg 550
   |||
1608  CTTCCACGCCAAGTCCCATACCAAGACGACCAACACACACCTCCACAGAGCA 1657
   |||
551  LysLeuUProglUProglUserArqGlyValValIleValAlaValIleVal 567
   |||
1658  AAGCTGCCGAGAGCCGAGAGCCGGGCTGTGTCATGTGGCTGTGATTGT 1707
   |||
567  lCysIleLeuValLeuAlaValLeuGlyAlaValAlaValLeuTrrPheLeuTrrL 584
   |||
1708  GTGATCTCTGTGCTGGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTA 1757
   |||
584  yAlaGlyGlyLysLeuProCysAArgAArgSerGlyLysGlnGlnIleThrLeu 600
   |||
1758  AGAAGAGCAAGCTTCCGTGTGAGCCCTCAGGAGAGAGAGAGATCACGGCTG 1807
   |||
601  ProProSerArqLysSerGlnLeuValValGlnValLysSerAspLysLe 617
   |||
1808  CCCCGCTGTCTGAAGACCCAGACTTGTATGTATGAATTAAGCAAGATTAAC 1857
   |||
617  uProglInglUmetGlyLeuLeuGlnGlnLysSerSerGlyAspLysAArgAla 634
   |||
1858  CCCAAGAAAGATGGGCTCTCTCTAGGAGCAGAGGGGAGAAAGAGAGGCTC 1907
   |||
634  roGlyAspGlnGlyGlnUtyrTrrLeuAspLeuAArgHis 646
   |||
1908  CCGGAGACCAAGGAGGAGAAATACATCATCTGAGAGCAT 1945
   |||

seq_name: gb_pr:U00MUC18B
seq_documentation_block: 3306 bp mRNA linear PRI 19-AUG-1994
LOCUS U00MUC18B
DEFINITION Human MUC18 glycoprotein mRNA, complete cds.
```


ACCESSION M28882
VERSION M28882.1 GI:529723
KEYWORDS MUC18 glycoprotein; cell adhesion molecule; immunoglobulin-like
sequence; integral membrane glycoprotein.
SOURCE Human cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3306)
AUTHORS Lehmann,J.M., Riethmuller,G. and Johnson,J.P.
TITLE MUC18, a marker of tumor progression in human melanoma, shows
sequence similarity to the neural cell adhesion molecules of the
immunoglobulin superfamily
Proc. Natl. Acad. Sci. U.S.A. 86 (24), 9891-9895 (1989)
JOURNAL MEDLINE 90093368
REFERENCE 2 (bases 1757 to 1945)
AUTHORS Sers,C., Kirsch,K., Rothbacher,U., Riethmuller,G. and Johnson,J.P.
TITLE Genomic organization of the melanoma-associated glycoprotein MUC18:
implications for the evolution of the immunoglobulin domains
Proc. Natl. Acad. Sci. U.S.A. 90 (18), 8514-8518 (1993)
JOURNAL MEDLINE 93391384
REFERENCE 3 (bases 1 to 3306)
AUTHORS Johnson,J.P.
TITLE Direct Submission
JOURNAL Submitted (06-OCT-1989) Judith P. Johnson, Institute for
Immunology, University of Munich, Goethestrasse 31, Munich, Germany
80336
COMMENT On Aug 15, 1994 this sequence version replaced gi:188861.
FEATURES
source
1..3306
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="zapyl, drop 1, drop 4, drop 4.2, drop 4.7, drop
4.11"
/cell_line="Mel Juso"
/tissue_type="melanoma"
8..91
8..1948
/codon_start=1
/product="MUC18 glycoprotein"
/protein_id="AAA20922.1"
/db_xref="GI:529724"
/translation="MGLPRLVCAFLAACCCPRVAVGPGEADQPAPELVEVEVGSTA
LAKGLSOGNLNVSHVDFVSHKEKRTLLPRVGGQSGSPGEVQSLSDGATLA
LTOVTPQDERFLPCGKRPSOEYRIQLRYKAPBEENIOVNPGLIVNKEPEEYAT
CVGNNGPIPIQVITVYKNGRPLKEKNVHTOSQTVSSGLYTLQSLKOLYKEDD
AQTCELNTRLPDSNHNKESRETVVYFYTEKYMLEVEVGHAKESDREYICLADG
NPPEHFIISKNPSTREAEETNDNGVLVEPARKHSERYECQANNDVLTSLISE
POELLVYSDVRYVSPAAPEROEGSLTLCESSDLEFQWLRETTDVLERGPLY
QLHLKREAGGGRVCAVSPISGLNRTQLKALIEGPPMAKEREKVMVKNMVLN
SCEASGHPRTISWNVNGTASEODOPQRLVSTLANVLTPELLETGEVCAASNDLGN
TSITFLVNTLTTPDSNTTGTSTASPHPTANSTERKLPEESGVYIVAVI
VCTIVLAVLCAVLEFYKKKRLPCRRSGKOETILPSPRKELVVEVAKSDLPREMGL
QGSQDKRAPEGDEKTIIDLRH"
92..1945
/product="MUC18 glycoprotein"
122..373
/note="immunoglobulin-like (V set) domain I"
misc_feature 467..694
/note="immunoglobulin-like (V set) domain II"
misc_feature 795..988
/note="immunoglobulin-like (C2 set) domain"
misc_feature 1076..1249
/note="immunoglobulin-like (C2 set) domain"
misc_feature 1337..1525
/note="immunoglobulin-like (C2 set) domain"
misc_difference 1766
/note="g is deleted"
misc_difference 1880-1881
/replace="g"
/citation="1"
misc_difference 1880-1881
/note="insert g"
/citation="1"

polYA_signal 3273..3278
BASE COUNT 799 a 916 c 924 g 667 t
ORIGIN
alignment_scores:
Quality: 3327.00 Length: 646
Ratio: 5.166 Gaps: 0
Percent Similarity: 99.690 Percent Identity: 98.916
alignment_block:
US-09-653-961-2 x HUMMUC18B ..
Align seg 1/1 to: HUMMUC18B from: 1 to: 3306
1 MGLGLYPRVPRARGLEUVALCYSAIAPHLEULEUVALAALACYSYCY 17
|||||
8 ATGGGGCTTCCAGAGCGTGGTCCGCTCTGCTGCGCGCTGCTG 57
|||||
17 SCYPRARGVALAAGLYVALPROGLYGLUALAGLUGINPROALAPROG 34
|||||
58 CTGTCTCCGCGTCCGCGGTGTCGCCGAGAGCGTGAAGCGCTG 107
|||||
34 LULEUVALGLUVALGLUVALGLYSERTHALALEULEULYSYSGLYLEU 50
|||||
108 AGCTGTGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 157
|||||
51 SERLINSENGINGLYASINLEUSERHISVALASPTIPHESEVALHISLY 67
|||||
158 TCCAGTCCCAAGGCAAGCTGACGATGTGACTGTTTCTGTCCACAA 207
|||||
67 SGLUYARGTHRLEULEPHEARGVALARGINGLYGLINGLINSERG 84
|||||
208 GGAGAGCGGAGCGCTCATCTTCGCTGTGCCAGGCGCAGGCGCAGAGCG 257
|||||
84 LUPROGLGLUTYRGLUGLARGLEUSERLEUGINASPARGGLYALATHR 100
|||||
258 AACCTGGGAGTACAGACAGCGGCTCAGCTCCAGGCAAGAGGCGCTACT 307
|||||
101 LEULALEUTHRGILNVALTHRPROGLINASPGLUNARGLIHPHELEUCYSGI 117
|||||
308 CTGGCCCTGACTCAAGTCACACCCCAAGACGAGCGCATCTTCTGTGCCA 357
|||||
117 NGLYLARGPROARGSERGLINGLYTYRARGILEGLINLEUARGVALTYRL 134
|||||
358 GGGCAAGCGCCCTCGTCCAGAGTACCGCATCCAGCTCGCGCTACAA 407
|||||
134 YSALAPROGLUGLUPROASNILEGLINVALASNPROLEUGLYLIPROVAL 150
|||||
408 AAGCTCCGAGAGCCCAACATCCAGGTCAACCCCTGGGCAATCCCTGTG 457
|||||
151 ASNSERTYSGIUPROGLUGLVALAALATHCYSAIAGLYARGASNGLYTY 167
|||||
458 AACACTAAGAGAGCCGAGAGAGTGCCTACCTGTAGGAGGAAGAGGCTA 507
|||||
167 TPRIILEPROGLINVALIETRPYLYRSASNGLYARGPROLEULYSGLUG 184
|||||
508 CCCCATTTCTCAAGTCATCTGTACAAAGAGGCGGCTCTGAAGAGAG 557
|||||
184 LULYSASNARGVALHISILEGINSESERGINTRVALIGLUSERSEGLY 200
|||||
558 AGAAGAACCGGCTCCACTTCAAGTCGTCGCCAGCTGTGAGTCAAGTGGT 607
|||||
201 LEUTYRTHRLEUGINSERILEULYSALAGINLEUVALLYSGIUAAPLY 217
|||||
608 TGTACACCTTGCAAGATATCTGAAGGACACAGCTGTGTAAGAGCAAA 657
|||||
217 SASPALAGINPHERYCYSGIULEUASNTRYRILEUROSERGLYASNH 234
|||||
658 ACATGCCAGTTTACTGTGAGCTCACTACCGGCTGCGCAGTGGAGACC 707
|||||
234 ISMELLYGLUSERARGGLUVALTHRVALPROVALPHETYPYPROTHRGLU 250
|||||

708 ACATGAGGAGTCCAGGAAATCACCCTCTTTTCTACCCGACAGAA 757
 251 LysValTrpLeuGluValGluProValGlyMetLeuLysGluGlyAsp 267
 758 AAGGTGGCTGGAACTGGAGCCCGTGGAAATGCTGAAGAGGGGACCG 807
 267 gValGluLeuArgCysLeuAlaAspGlyAsnProProHisPheSerI 284
 808 CGTGAATATCAGGTGTGTGGTATGATGCNAACCTCCACACACTTCAGA 857
 284 LeSerLysIleAsnProSerThrArgGluAlaGluGluThrThrAsn 300
 858 TCAGCAAGAGAAACCCACACAGGAGGAGGAGAAAGAACCAACCAAC 907
 301 AspAsnGlyValLeuValLeuGluProAlaArgLysGluHisSerGly 317
 908 GACACGGGGCTCTGGTGTGGAGCTGCCCGGAAGAACACAGTGGCG 957
 317 gTyrGluCysGlnGlyLeuAspLeuAspThrMetIleSerLeuLeuSer 334
 958 CTATGATGTACAGCCCTGGAACTGGACACCATGATTCGTCTGAGTG 1007
 334 LProGluGluLeuLeuValLeuValLeuValLeuValLeuValLeu 350
 1008 AACCAAGAACTACTGGTGAATGATGTCTGACGTCCGAGTGAATCC 1057
 351 AlaAlaProGluArgGlnGluGlySerSerLeuThrLeuThrCysGlu 367
 1058 GCAGCCCCCTGAGAGACAGGAGGACGACCCCTGACCTGTGAGAGC 1107
 367 LeuLeuSerSerGlnAspLeuGluPheGluTrpLeuArgGluThrGly 384
 1108 ACAGAGTACCCAGGACCTGAGTCTCACTGCTGACAGAGAGACAGCC 1157
 384 LValLeuGluArgGlyProValLeuGluIleHisAspLeuLysArg 400
 1158 AGCTCTGGAAGAGGGGCTGTGCTTCACTGCTGATGACCTGAAGGGAG 1207
 401 AlaGlyGlyGlyTyrArgCysValAlaSerValProSerIleProGly 417
 1208 GCAGAGGCGGCTATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1257
 417 uAsnArgThrGlnLeuValAsnValAlaIlePheGlyProProTrpMet 434
 1258 GAACCGCACACAGCTGTGTCAGCTGGCCATTTTGGCCCCCTTGAAGG 1307
 434 LaphelLysGluArgLysValTrpValLysGluAsnMetValLeuAsn 450
 1308 CATTCAGAGAGAGAGGTGTGGGTGAAGAAATATGTGTGATCTG 1357
 451 SerCysGluAlaSerGlyHisProArgProThrIleSerTrpAsnVal 467
 1358 TCTGTGTGAAGGTCAAGGACCCCGGCGCCACCATCTCTTGGAACTCA 1407
 467 nGlyThrAlaSerGluGlnAspGlnAspProGlnArgValLeuSerThr 484
 1408 CGGCGCGCAAGTGAACAAACCAAGATCCACAGGAGTCTCTGAGCACC 1457
 484 euAsnValLeuValThrProGluLeuLeuGluThrGlyValGluCysThr 500
 1458 TGAATGTCTCTGTGACCCCGAGCTGTGGAGACAGGTGTGATGACAG 1507
 501 AlaSerAsnAspLeuGlyLysAsnThrSerIleLeuPheLeuGluLeu 517
 1508 GCCCTCAACGAGCTGTGGCAAAAACACAGCATCTCTTCCGAGAGCTGT 1557
 517 LAsnLeuThrThrLeuThrProAspSerAsnThrThrThrGlyLeuSer 534
 1558 CAATTTAAACCCCTCACACACAGACTCCAAACACACACATGGGCTCAGCA 1607
 534 hrSerThrAlaSerProHisThrArgAlaAsnSerThrSerThrGluArg 550
 1608 CTTCACATGCGCAGTCTCATACAGAGCAACACACACTCTCCACAGAGAGA 1657

551 LysLeuProGluProGluSerArgGlyValValIleValAlaValIle 567
 1658 AAGCTCCGCGAGCCGGAGAGCCGGCGGTGTGTATCTGCTGTGATGT 1707
 567 LysIleLeuValLeuAlaValLeuGlyAlaValLeuTyr-PheLeuTyr 584
 1708 GTGCATCTGTCTCTGGCGGTGTGTGGCGCTGTCTTATTTCTCTATA 1757
 584 yLysGlyLysLeuProCysArgArgSerGlyLysGlnIleThrLeu 600
 1758 AGAAGGCGCAAGCTCCGTGCAGCGCTCAGGAGAGAGATACAGCTG 1807
 601 ProProSerArgLysSerGluLeuValValGluValLysSerAspLys 617
 1808 CCCCCTCTCTTAAGACCGAATGATGTAAGTAAATCAATTAAGCT 1857
 617 uProGluGluMetGlyLeuLeuGlnGlnGlySerSerGlyAspLysArg 634
 1858 CCCAGAGAGATGGCGCTCCCTGCAGGCGCACAGCGGTGACAAAGGCTC 1907
 634 roGlyAspGlnGlyGluLysTyrIleAspLeuArgHis 646
 1908 CGGGAGACCAAGGAGAGAAATACATCATCTGAGGAT 1945

seq_name: gb_ro:AB035508

seq_documentation_block:

LOCUS AB035508 2890 bp mRNA linear ROD 03-Oct-2000

DEFINITION Mus musculus mRNA for 1-glicerol/MUC18, complete cds.

ACCESSION AB035508

VERSION AB035508.1 GI:10566952

KEYWORDS 1-glycerol/MUC18.

SOURCE Mus musculus (strain: BALB/c) male cDNA to mRNA.

ORGANISM Mus musculus (strain: BALB/c) male cDNA to mRNA.

REFERENCE 1 (bases 1 to 2890)

AUTHORS Taira, E., Okumura, S. and Miki, N.

JOURNAL Published Only in Database (2000) In press

REFERENCE 2 (bases 1 to 2890)

AUTHORS Taira, E., Okumura, S. and Miki, N.

JOURNAL Direct Submission

TITLE Submitted (03-Dec-1999) Eiichi Taira, Osaka University Medical

School, Department of Pharmacology, 2-2 Yamada-oka, Suita, Osaka

565-0871, Japan (E-mail: etaira@pharm.med.osaka-u.ac.jp,

Tel.: +81-6-6879-3521, Fax: +81-6-6879-3521)

FEATURES Location/Qualifiers

1..2890

/organism="Mus musculus"

/strain="BALB/c"

/db_xref="taxon:10090"

/sex="male"

20..1966

/gene="1-glicerol/MUC18"

/product="1-glicerol/MUC18"

/codon_start=1

/protein_id="BAB16050.1"

/translation="MGLPKLVCFVLAFAACCCRRAGVPEEKOPVTPDIVEAGVS

TALCKGSRASNGFSDWFLIKERQILIFRHGKGGRREYERHRSLODSVAT

LALSHVPRHBRMFLCKSKRPLODHYELVEVPAKVVVEVPGGLKEGDHVRILT

ATCGRNGRTPROYLWKNSPLDDEENRHIOSSQTVESGLITLKSIVLSARLVED

KDAQFCELSLRPSGNHMKESKEVTVPAKVVVEVPGGLKEGDHVRILT

DGNQPHPTINKKDPSTGEMEBESTDENGILSEPAKHSGLYOCOSIDLETTITLS

SDPLELVNVYSDVQNPATADEVQESLITLCEAESNODLEPEMLRDKTGOLGEL

VLOLNVRREAGGRYLCMASVPRVGLNRTOLVSGIFGSPMALEKRWKQENAVL

NLSGEASGHOPTISMVNGSATENMPDPTQVSTLVNLTTPPLLTGACACTSNLSL

SNITTYLKLVTITLTLLPDSQTTGLSTLSTVSHTRANSTSTKRLPQPSKGVIVA

VIVCTILAVLVGAALVFLYRKGLPCGSRSGKOBITLPPRKSEFVEVSKDLPEEMA


```

549 gluArgLysLeuProGluProGluSerArgGlyValValIleValAlaIaIa 565
|||||
1670 GAGAAAGAGCTGCCACAGCAGAGCAAGAGTGTGTCATCTGCGCTGT 1719
565 IILeValGlyIleLeuValIleAlaIaValLeuGlyValAlaValLeuTyrPhe 582
|||||
1720 GATAGTGTGTACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1769
582 euTyrLysGlyLysLeuProCysArgArgSerGlyLysGlnGluIle 598
|||||
1770 TCTACAAGAGGCGCAAGCTGCATCTGCACGCTCGGAAACAGAGATC 1819
599 ThrLeuProProSerArgLysSerLeuValValIleGluValLysSerAs 615
|||||
1820 ACGCTGCCCCCGACCTCGTAAGTGAATTTGTAAGTTAACTGACAGA 1869
615 pLysLeuProGluGluMetGlyLeuLeuGlnGlySerSerGlyAspLysA 632
|||||
1870 TAAAGTCCCAAGAGATGCTCTCTCTCAGCGCAGCAACGCTGACAGA 1919
632 tGAlaProGlyAspGlnGlyLysTyrIleAspLeuArgHis 646
|||||
1920 GGGCTCCAGAGACAGGAGAGAAATACATCGATCTGAGCAT 1963
seq_name: gb_ro:AB035506

seq_documentation_block:
LOCUS AB035506 2106 bp mRNA linear ROD 03-OCT-2000
DEFINITION Rattus norvegicus mRNA for l-glicerIn/MUC18, complete cds.
ACCESSION AB035506
VERSION AB035506.1 GI:10566948
KEYWORDS l-glicerIn/MUC18.
SOURCE Rattus norvegicus (strain: Sprague-Dawley) male cDNA to mRNA.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 2106)
AUTHORS Taira,E., Okumura,S. and Miki,N.
TITLE Functional analysis of rat glicerIn/MUC18
JOURNML Published Only in Database (2000) In press
REFERENCE 2 (bases 1 to 2106)
AUTHORS Taira,E., Okumura,S. and Miki,N.
TITLE Direct Submission
JOURNML Submitted (03-DEC-1999) Eiichi Taira, Osaka University Medical
School, Department of Pharmacology, 2-2 Yamada-oka, Suita, Osaka
565-0871, Japan (E-mail:etaira@pharm.med.osaka-u.ac.jp,
Tel:+81-6-6879-3521, Fax:+81-6-6879-3521)
FEATURES
source
location/Qualifiers
1..2106
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/sex="male"
17..1963
/gene="l-glicerIn/MUC18"
17..1963
/gene="l-glicerIn/MUC18"
/codon_start=1
/product="l-glicerIn/MUC18"
/protein_id="BAB16048.1"
/db_xref="GI:10566948"
/translation="MGLPRLYCAFLFAACCCRSATGVPEEKPRTTPDPEVEVGN
TALCKGPAHSGNMFQSEWFLIKERQIPIERHNGKGSPEPEYHRISLSPGAT
LALSGVPRHDMFLCKSKDRPDHLYOQVYKAREPTQAVLGIHVDIQLKEV
ATCVGRNGYRPIQYIWKNGRPRLOEENRHHIQSQTVESGLITLKSYSARVYKD
KDAQFYCEVSYPSPGNMKSKESEVTVPVLPARKVVEVEPVGLKEGSHVIRCLT
DGNPOPHFTINKKNPSTEMEESTDEGSLSTEPRAKHHSYVOCOSLDLETVMIS
SDPLFLNVYSDVDPDTABVEQDSILTLTCAEENODLEFEMLDKLGOLITGKPS
ILOLNMYKREAGRGYLCVASYSPGLNRRPRVYGLGSPMAAKERKVMWAQENAL
NLSCEASGHPPTISWNINGSATERNPDPRQYVSTLVNLVLTPELLTGACTASNSIG
SYTVIILKLTITLTTHDSSQTGLSTPVSFHSRANSTSEKKLPQDSKGVYIA

```

```

VIVCTLVAVLAVGATLYFYFKKGLPCGRSGSKQEIITLPTPRKSEFVEVKSCLKPEEMA
LLQSGNGKRRAPGDGEEKYIDLRRH"
BASE COUNT 536 a 573 c 587 g 410 t
ORIGIN
alignment_scores:
Quality: 2587.00 Length: 648
Ratio: 4.297 Gaps: 1
Percent Similarity: 92.901 Percent Identity: 74.537
alignment_block:
US-09-653-961-2 x AB035506 ..
Align seg 1/1 to: AB035506 from: 1 to: 2106
1 MetGlyLeuProArgLeuValCysAlaPheLeuLeuAlaIleCysCys 17
|||||
17 ATGGGGCTGCCAGGCGTGTGCTGCGCTTCTTATTTGCTGCTGCTGCTG 66
17 SCYSProArgValAlaGlyValProGlyGluAlaGln.....ProA 32
|||||
67 CTGTGCGACGCCCGCGGTGTGCCAGAGAGAAACAGCCACACCA 116
|||||
117 CGCCGACCCGCGGTGAGGTGAAGTGGCAACACGCGCTTCTGAAGT 166
32 lAProGluLeuValGluValGluValGlySerThrAlaLeuLysCys 48
|||||
49 GlyLeuSerGlnSerGlnGlyAsnLeuSerHisValAspIrrPheSerA 65
|||||
167 GCGCCCGCACATCCCTCAGGCACTTCCAGCAAGTGAAGTTTGTGAT 216
167 IHisGlyLysArgThrLeuIlePheArgValArgGlnGlyGln 82
|||||
217 TCACAGAGAGAGGACATACCATTTCCGTGTGCACAGGCAAGGCC 266
|||||
82 lnsErGluProGlyGluTyrGlnArgLeuSerLeuGlnAspArgGly 98
|||||
267 AGAGTGAACCTGGGGAATGATGACACCGCTCAGCCGCCAGCA 316
99 AlathrLeuAlaLeuThrGlnValThrProGlnAspGluArgIlePhe 115
|||||
317 GCTACTGTGGCTGAGTCAAGTACTCTCCATGATATACCGGATGTTCT 366
115 uCysGlnGlyLysArgProArgSerGlnGlyTyrArgIleGlnLeuArg 132
|||||
367 GTGTAAGAGCAAGCAGCAGCGCCAGATCACTATGTTCAGGTGCAAG 416
132 alTyrLysAlaProGluGluProAsnIleGlnValAsnProLeuGlyIle 148
|||||
417 TCTACAAGCTCCAGAGAGCAACTATCCAGGCCAGCTGTGGCATC 466
149 ProValAsnSerLysGluProGluGluValAlaIaThrCysValGlyArg 165
|||||
467 CACGTGAGCATACAAAGAGCTTAAGAGGTGTCTGCTGGGGAGAA 516
165 nGlyTyrProIleProGlnValIleTyrTyrLysAsnGlyArgProLeu 182
|||||
517 TGGTACCCCATTCCTCAAGTCATCTGTACAAAGAGCGCGGCTTGC 566
182 YSGlGluLysAsnArgValHisIleGlnSerSerGlnThrValGlnSer 198
|||||
567 AAGAGAGAGAGAACCGCTTATATCCAGTGTACACAGCCGTGATCC 616
199 SerGlyLeuTyrThrLeuGlnSerIleLeuLysAlaGlnLeuValLysG 215
|||||
617 AGTGCTTGTAACCTTGAAGAGTGTCTAGTACAGGGATGCAAGAA 666
215 uAspLysAspAlaGlnPheTyrCysGluLeuAsnTyrArgLeuProSerG 232
|||||
667 AGACAAAGATGCCCGAGTTTACTGTGAAGTCAGCTACCGCGGCCAGTG 716
232 lYAsnHisMetLysLeuSerArgGluValThrValProValPheTyrPro 248

```

|||||.....
717 GGAACCGCATGAAGGAATCTAAGAGGCTGCTGCTCTTATACCT 766
249 ThrGluValTTrpLeuGluValAlaGluProValGlyMetLeuValGlu 265
767 GCAGAAAAAGTGTGGGTGAGGTAGAACCCCTGGGACTGCTGAAGGAG 816
265 yAspArgValGluLeuArgCysLeuAlaAspGlyAsnProProHisP 282
817 GGACCATGTGAAGATCAGGTGCTGTAGTACGAGGCAACCTCAACCCACT 866
282 hSerIleSerIleGlnAsnProSerThrArgGluAlaGluGluThr 298
867 TCACATATCAACAAGAAAGAACCCAGCTGAGATGGAAGAGAGAGC 916
299 ThrAsnAspAsnGlyValLeuValLeuGluProAlaArgLysGluHis 315
917 ACCGATGAAAAAGCTCTCTCTCTGAGCCCGCCAGAAAGCCACACAG 966
315 rGlyArgTyrGluCysGlnGlyLeuAspLeuAspThrMetIleSerLeu 332
967 TGGGTGTACCAATGTACAGGCTGACTGTGAACACTACAGTTATGTGT 1016
332 euSerGluProGlnGluLeuValAsnTyrValSerAspValArgVal 348
1017 CGAGTGACCCCTGAGCTGTGTGAACTATGTCTGATGTTCAAGTG 1066
349 SerProAlaAlaProGluArgGlnGluGlySerSerLeuThrLeuThr 365
1067 GATCCAACTGCCCCCTGAAGCTCCAGAGAGGTATAGCTCACACTGACTG 1116
365 sGluAlaGluSerSerGlnAspLeuGluPheGlnTrpLeuArgGluThr 382
1117 TAAGCAGAGAGTAAACCGACCTTGAGTTGAGTGCTGAGAGCAAGA 1166
382 hArgGluValLeuGluArgGlyProValLeuGlnLeuHisAspLeuLys 398
1167 CAGGCACTCTTGTGGAAAGGGGCCCACTTCACTCAACACAGTGAGA 1216
399 ArgGluAlaGlyGlyTyrArgCysValAlaSerValProSerIlePr 415
1217 AGGAGAGAGGGGAGAGATATCTCTGTGCACTGCTCCACAGCTTCC 1266
415 oGlyLeuAsnArgThrGlnLeuValAsnValAlaIlePheGlyProTr 432
1267 TGGCTTGAATGCACACGCGGGTCAGCGTGCGATTTTGGGCCCGCT 1316
432 rPheAlaPheLysGluArgLysValTTrpValLysGluAsnMetValLeu 448
1317 GGATGGCAGCAAGAGAGAGAGGTGGCGCAAGAAATGCAATGCTG 1366
449 AsnLeuSerCysGluAlaSerGlyHisProArgProThrIleSerTrpAs 465
1367 AATGTGTTTGTGAGGCTTCAGGCCATCTCAAGCCACCACTCTCTGGA 1416
465 nValAsnGlyThrAlaSerGluGlnAspGlnAspProGlnArgValLeu 482
1417 TATCAATGGCTCGCACTGAATGAACCTGATCCTCAGACAGTAGAGA 1466
482 eTrpLeuAsnValLeuValThrProGluLeuLeuGluThrGlyValGlu 498
1467 GCACCTTGAATGCTTGTGACCCAGCTTCGGAACCGGTCAGAGA 1516
499 CysThrAlaSerAsnAspLeuGlyLysAsnThrSerIleLeuPheLeuGlu 515
1517 TGTACAGCTCACTCCCTGGGCTCATACACACAGGCTCATCTTGA 1566
515 uLeuValAsnLeuThrThrLeuThrProAspSerAsnThrThrArgLys 532
1567 GCTGCTCACTTAAACCACTCACATGACTCCAGCAACCACTGGCC 1616
532 euSerThrSerThrAlaSerProHisThrArgAlaAsnSerThrSerThr 548
|||||.....

1617 TCAGCACCCCAACGATCAGTCTCTACAGCAGAGCCAAACAGCACTCCACA 1666
549 GluArgLysLeuProGluSerArgGlyValValIleValAlaVala 565
1667 GAGAAAAAGTGTCCACACAGCAGGAGAGACAAAGTGTGTATGTGCTGT 1716
565 lIleValCysIleLeuValLeuAlaValLeuGlyAlaValLeuTyrPheL 582
1717 GATAGTATGACCTGTGCTGCTGCTGAGGCGCTACCTCTATTA 1766
582 euTyrIleLysGlyLysLeuProCysArgArgSerGlyLysGlnGlu 598
1767 TCACAAAGAAAGGCAACCTGCTGTGAGCGCTCAGAGAAACAGGAGATC 1816
599 ThrLeuProProSerArgLysSerGluLeuValAlaGluValLysSer 615
1817 ACGGTGCCCCGAGCTGTAAGACTGAATTTGTATGTTGAAGTTAAGTCACA 1866
615 pLysLeuProGluGluMetGlyLeuLeuGlnGlySerSerGlyAspLys 632
1867 TAAGCTCCAGAGAGATGCTGCTCTCCTCAGGCGCAGCAACGGTGACAGA 1916
632 rGluAlaProGlyAspGlnGlyGluLysTyrIleAspLeuArgHis 646
1917 GGGCTCCAGAGACAGAGGAGAGAAATACATCATCTGAGGAT 1960
seq_name: gb_ro:AB035509

seq_documentation_block:

LOCUS AB035509 2772 bp mRNA linear ROD 03-Oct-2000
DEFINITION Mus musculus mRNA for s-gicerin/MUC18, complete cds.

ACCESSION AB035509
VERSION AB035509.1 GI:10566954

KEYWORDS s-gicerin/MUC18.
SOURCE Mus musculus (strain: BALB/c) male cDNA to mRNA.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 2772)
Taira, E., Okumura, S. and Miki, N.

TITLE mouse s-gicerin/MUC18
JOURNAL Published Only in Database (2000) In press

REFERENCE 2 (bases 1 to 2772)
Taira, E., Okumura, S. and Miki, N.

AUTHORS Direct Submission
TITLE Submitted (03-Dec-1999) Eiichi Taira, Osaka University Medical
School, Department of Pharmacology; 2-2 Yamada-oka, Suita, Osaka

565-0871, Japan (E-mail: etaira@pharmal.med.osaka-u.ac.jp,
Tel: +81-6-6879-3521, Fax: +81-6-6879-3521)

FEATURES
Location/Qualifiers

1..2772
/organism="Mus musculus"
/strain="BALB/c"

/db_xref="taxon:10090"
/sex="male"

20..1840
/gene="s-gicerin/MUC18"

20..1840
/gene="s-gicerin/MUC18"

20..1840
/gene="s-gicerin/MUC18"

20..1840
/gene="s-gicerin/MUC18"

20..1840
/gene="s-gicerin/MUC18"

20..1840
/gene="s-gicerin/MUC18"

20..1840
/gene="s-gicerin/MUC18"

20..1840
/gene="s-gicerin/MUC18"

20..1840
/gene="s-gicerin/MUC18"

20..1840
/gene="s-gicerin/MUC18"

20..1840
/gene="s-gicerin/MUC18"

20..1840
/gene="s-gicerin/MUC18"

BASE COUNT 675 a 706 c 749 g 642 t
ORIGIN
alignment_scores: Quality: 2415.50 Length: 648
Ratio: 4.275 Gaps: 2
Percent Similarity: 87.191 Percent Identity: 70.833
alignment_block:
US-09-653-961-2 x AB035509 ..
Align seg 1/1 to: AB035509 from: 1 to: 2772
1 MetGlyLeuProArgValAlaGlyValAlaPheLeuAlaAlaCysCysCys 17
|||||
20 ATGGGGCTGCCAAACTGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 69
17 scysProArgValAlaGlyValAlaPheLeuAlaAlaCysCysCys 32
|||||
70 CTGTGCGCGCGCGCGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCG 119
32 laProGluLeuValGluValGluValGlySerThrAlaLeuLeuLeuLeu 48
:|||||
120 CGCCCGACCGTGTGGAGCGCAAGTGGCGACACCGCTTCACAGTGT 169
49 GlyLeuSerGlnSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 65
|||||
170 GGGCCCTCAGGGCTCAGGCACTTACAGCAAGTGGAGCTGGTGTGTGAT 219
65 lHisLysGluLysArgThrLeuLeuPhePheArgValAlaArgGlnGlnGly 82
:|||||
220 TCACAAAGAGAGGACAGTACTGATTTCCGTGTCGACCAAGCAAGCGCC 269
82 lnsertLuproGluLutryGluGlnArgLeuSerLeuLnsparGly 98
|||
270 AGCGGAACTGGTGAATATGAGCACCGGCTTACCTCCAAAGCTCGGTG 319
99 AlaThrLeuAlaLeuThrGlnValAlaThrProGlnAspGluArgIlePheLe 115
|||||
320 GCTACTCTGGCCCTGAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTG 369
115 ucysGlnGlyLysArgProArgSerGlnLutryArgIleGlnLeuArgV 132
|||||
370 GTGTAGAGAGCAAGCGACCGGCTCCAGGATCAGTGTGAGCTTCAGG 419
132 alLutryLysAlaProGluLuproAsnIleGlnValAsnProLeuGlyIle 148
|||||
420 TCTTCAAAGCCCGAGAGAACCAATATTCACAGCAATGTCTGGGGCATC 469
149 ProValAsnSerLysGluProGluGluValAlaThrCysValGlyArgAs 165
|||||
470 CATGTGAGAGCGCAAGAGCTCAAGAGGTTGCTACTGCTGGGGAGAA 519
165 nGlyLutryProIleProGlnValIleTpyLysAsnGlyArgProLeuL 182
|||||
520 CGGCTACCCCATTCCTCAAGCTCTATGTAACAAGACAGTCTGCCCTGC 569
182 ysgLutLutLysAsnArgValHisIleGlnSerSerGlnThrValGluSer 198
:|||||
570 AAGAGGAGGAGAACCGAGTTCAATATCCAGTCAATGTCGAGTCC 619
199 SerGlyLeuLutryThrLeuGlnSerIleLeuLysAlaGlnLeuValLysG 215
|||||
620 AGTGCTGTGACACCTTGAAGAGTGTCTGAGTGCACCGCTAGTTAAGA 669
215 uaspyLysAspAlaGlnPheTyrCysGluLeuAsnTyrArgLeuProSerG 232
|||||
670 AGACAAAGATGCCAGTTTACTGTGAAGTCAAGCTACGCGTACCAAGG 719
232 lLysnHisMetLysGluSerArgLutValThrValProValPheTyrPro 248
|||||

720 GGAACCATATGAAGGAATCTAAGAGGTCACTGCTCCTGTTTCTACCT 769
249 ThrGluLysValITrPheLutValGluProValGlyMetLeuLysGluL 265
:|||||
770 GCAGAAAAAGTGTGGGTGAGGTAGAGCTTGGGGCTGTGAAGGAAGG 819
265 yAspArgValGluLeaArgCysLeuAlaAspGlyAsnProProHisP 282
|||||
820 GGATCATGTGACAAATCAGGTGCTGACAGATGGCAACCTCAACCCACT 869
282 heserLleSerLysGlnAsnProSerThrArgGluAlaGluGlnThr 298
:|||||
870 TCACATATCAACAGAGAGGCCAGCTGGGAGATGAGAGAGAGAGC 919
299 ThrAsnAspAsnGlyValLeuValLeuGluProAlaArgLysGluHisSe 315
:|||||
920 ACCGATGAATAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 969
315 rGlyArgTyrGluCysGlnGlyLeuAspLeuAspThrMetLleSerLeuL 332
|||||
970 CGGGCTTACCAAGTGTAGAGTGTGAGCTGGAACATCACTACACTGT 1019
332 euSerGluProGlnGluLeuLeuValAsnTyrValSerAspValArgVal 348
|||||
1020 CAAGTGAACCCCTGAGCTGCTGTGAACTATGTGTGATGATTCACAGT 1069
349 SerProAlaAlaProGluLutArgGlnGluLysSerSerLeuThrLeuTrpCy 365
:|||||
1070 AATCCAACTGCCCTCGAAGTCCAGGAAGTGAAGCTCAACGCTGACTG 1119
365 sGluAlaGluSerSerGlnAspLeuGluPheGlnTrpLeuArgGluLut 382
|||||
1120 CGAGGCAAGAAAGTAAACCGAGACTTGAATGATGCTGAGAGACAGAA 1169
382 rGluGluValLeuGluLutArgLutProValLeuGlnLeuHisAspLeuL 398
|||||
1170 CAGGCACTGCTGGAAAGGCTCCGCTCCAGCTTCAACAAAGCTGAGA 1219
399 ArgGluAlaGlyGlyLutYrArgCysValAlaSerValProserIlePr 415
|||||
1220 CGGGAAGCAGGGGGAGCGTATCTGCAATGCGATCTGCCCAAGATGCC 1269
415 oGlyLeuAsnArgThrGlnLeuValAsnValAlaIlePheGlyProProt 432
|||||
1270 TGGCTTGAATGTCGCCAGCTGCTGAGCTGGCGATTTTGGGCCCAT 1319
432 rPheAlaPheLysGluArgLysValITrPValLysGluAsnMetValLeu 448
|||||
1320 GATGGCATTTAAAGAGAGAGAGAGGTGTGGTGCAGAGAAATGCAGTCTG 1369
449 AsnLeuSerCysGluAlaSerGlyHisProArgProThrIleSerTrpAs 465
|||||
1370 AATCTGCTTGTGAGGCTTCAAGACATCTCAGCCCAACATCTCTTGAA 1419
465 nValAsnGlyThrAlaSerGluGlnAspGlnAspProGlnArgValLeuS 482
|||||
1420 TGTCAATGTTCGGCACTGAATGAACCCAGATCCACAGACAGTATGTA 1469
482 erThrLeuAsnValLeuValITrProGluLeuLeuGluITrGlyValGlu 498
|||||
1470 GCACCTTAATGTCTTGTGACGCCAGAGCTTCTGGAGACAGGTGCAGAG 1519
499 CysThrAlaSerAsnAspLeuGlyLysAsnThrSerIleLeuPheLeuG 515
|||||
1520 TGTACAGCTTCAACTCTCTGCTGCTCAAAACACACACACATGTTTCGAA 1569
515 uLeuValAsnLeuThrITrLeuThrProAspSerAsnITrThrGlyL 532
:|||||
1570 GCTGTCACTTAAACCACTTCATACCTGACTCCAGCAACCACTGACC 1619
532 euSerITrSerITrAlaSerProHisITrArgAlaAsnSerITrSerITr 548
|||||
1620 TCAGCACCTTCAAGTCAAGTCTCAGACCAAGCAAGCAAGCAAGCAAG 1669

549 GUAArgLysLeuProGluSerArgGlyValValIleValAlaVala 565
|||||
1670 GAGAAAGAGCCACACCCAGAGAGCAAGGTGGTGCATCGGCTGT 1719
565 ILeValCysIleLeuValIleAlaValLeuGlyAlaValLeuTyrPheL 582
|||||
1720 GATATGTGTACCTTGTGCTGTGGTGTCTGGTCTCTCATTTCC 1769
582 eUTYrLYsGlyLysLeuProCysArgArgSerGlyLYsGlnGluIle 598
|||||
1770 TCTACAAAGAGGCGCAAGCTGCATGTGAGCGTGGGAAACAGAGAT. 1818
599 ThrLeuProProSerArgLYsSerGluLeuValValGluValLYsSerAs 615
1818 1818
615 pLYsLeuProGluGluMetGlyLeuGlnGlySerSerGlyAspLYsA 632
1818 1818
632 rGAlaProGlyAspGlnGlyLysTyrIleAspLeuArgHis 646
1819GGAGAGAAATACATCATGTGAGGCAT 1845
seq_name: gb_ro:AB035507
seq_documentation_block: 1988 bp mRNA linear ROD 03-OCT-2000.
LOCUS AB035507
DEFINITION Rattus norvegicus mRNA for s-glicerol/MUC18, complete cds.
ACCESSION AB035507
VERSION AB035507.1 GI:10566950
KEYWORDS s-glicerol/MUC18
SOURCE Rattus norvegicus (strain:Sprague-Dawley) male heart cDNA to mRNA.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1988)
AUTHORS Taira,E., Okumura,S. and Miki,N.
TITLE rat s-glicerol/MUC18
JOURNAL Published Only in Database (2000) In press
REFERENCE 2 (bases 1 to 1988)
AUTHORS Taira,E., Okumura,S. and Miki,N.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-1999) Eiichi Taira, Osaka University Medical
School, Department of Pharmacology; 2-2 Yamada-oka, Suita, Osaka
565-0871, Japan (E-mail:etaira@pharm.med.osaka-u.ac.jp,
Tel:+81-6-6879-3521, Fax:+81-6-6879-3521)
FEATURES
source
1. 1988
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/sex="male"
/tissue_type="heart"
17. 1837
/gene="s-glicerol/MUC18"
17. 1837
/gene="s-glicerol/MUC18"
/codon_start=1
/product="s-glicerol/MUC18"
/protein_id="PAB16049.1"
/db_xref="GI:10566951"
/translation="MGCLPRVCAFLFACCCCSATGVPGEKQPTPTDPVVEVGN
TALKCGPAHSGNFSQVEMFLHKEKQPTIFRHQKSGSEFEBHRLSLGPGAT
LALSQVPHDRMFLKSKQRPQDHYVQLOVYKAPPEPTIQANVLGIHVDIQLKELV
ATCGANGYIPQVITWKNGRPLQDEENRVLIOSSQVSSGLYTLKSIVLSARVVED
KDAQFEVSYRPPSGNRKESKEVTVPLVPAKVVEVPGVLKEDGKVRICVET
DGNQPHFTINKKRSPTSEEMSESTDENGVLSLEPAKHSGYVCOOSDLETTVMIS
SDPLEILYVSDVQVPTAPVQEGSLTLTKAKENQDLERFEMLEDRKQGLGKGP
ILQLNVKREKGRYLCVASPSVPGILNTRIRVSVGLFGSPMAAKERKRWADENAL
NLSCASGHPDPTISNINSGATSEWNPQTVSTLNLVTPLELLETGACBTASNSLG

SYTIVILIKVTLTTLTHDSQTTGLSTPTVPSHRANSTSTKRLPODESKGVIVA
VIVCTVLAVALVATLYEYFKKGLPCGRSGKQEMERTSI"
BASE COUNT 503 a 544 c 553 g 388 t
ORIGIN
alignment_scores:
Quality: 2360.50 Length: 648
Ratio: 4.193 Gaps: 2
Percent Similarity: 86.883 Percent Identity: 68.981
alignment_block:
us-09-653-961-2 x AB035507 ..
Align seg 1/1 to: AB035507 from: 1 to: 1988
1 MetGlyLeuProArgLeuValCysAlaPheLeuValAlaLacCysCys 17
|||||
17 ATGGGCGTCCCGCAGCGTGTGCTGCTCTTATTCGCTGCTGCTGCTG 66
17 sCYsProArgValAlaGlyValProGlyGluAlaGln.....ProA 32
|||||
67 CTGTGCGACGCCAGCGGTGTGCGAGAGAGAAAGCAGCCACACCA 116
32 LaproGluLeuValGluValGlySerThrAlaLeuLeuLYsCys 48
|||||
117 CGCGCGACCGCGGTGAGGTAGAAAGTGCGCACACGCGCTTCGAAGT 166
49 GlyLeuSerGlnSerGlnGlyAsnLeuSerHisValAspTyrPheSer 65
|||||
167 GCGCCCGCACATCCCTCAGCACTTCAGCAAGTGGAATGTTTGTAT 216
65 IHisLYsGlyLysArgThrLeuPheArgValArgGlnGlyGln 82
|||||
217 TCACAAAGAGAGGAGATACCATTTTCCGTGTGCACACAGGAGAGGCC 266
82 InSerGluProGlyGlyTyrGlnGlnArgLeuSerLeuGlnAspArgGly 98
|||||
267 AGAGTGAACCTGGGGAATATGAACACCGCTCAGCCCTCCAGCGCCACGA 316
99 AlaThrLeuAlaLeuThrGlnValThrProGlnAspGluArgLLePheLe 115
|||||
317 GCTACTCTGGCCCTGAGTCAAGTCACTCCCATGATGATGCGGATGTTCC 366
115 uCYsGlnGlyLysArgProArgSerGlnGlyTyrArgGlnLeuArg 132
|||||
367 GTTAAGAGCAGACGACGACGCGCCAGATCATTGTTCTACCTGCAG 416
132 alTYrLYsAlaProGluGluProAsnIleGlnValAsnProLeuGlyIle 148
|||||
417 TCTACAGGCTCCAGAGAGACCACTATCCAAAGCAGCGTGGGCATC 466
149 ProValAsnSerLYsGluProGluGlnValAlaThrCysValGlyArgAs 165
|||||
467 CACGTGACATACAAAGAGCTTAAGAGAGTTGCTACCTGTGGGAGAA 516
165 nGlyTYrProIleProGlnValIleTyrLYsAsnGlyArgProLeu 182
|||||
517 TGGCTACCCCATTCCTCAAGCATCTGTACAAAGAAGCGCGCCCTGTC 566
182 ySGluGlyLysAsnArgValHisIleGlnSerSerGlnThrValGluSer 198
|||||
567 AAGAGAGAGAGAACCGCGTTCATATCCAGTGTACAGACCGTGGAGTCC 616
199 SerGlyLeuTyrThrLeuGlnSerIleLeuLYsAlaIleuValLYsGly 215
|||||
617 AGTGGCTGTACACCTGAAGAGTGTCTAAGTCAAGGAGGTAGTCAAGGA 666
215 uAspLYsAspAlaGlnPheTyrCysGluLeuAsnTyrArgLeuProSer 232
|||||
667 AGACAAAGATGCCAGTTTACTGTGAAGTCAGCTACCGGCGGCGACGT 716
232 IAsnHisMetLYsGluSerArgGluValThrValProValPheTyrPro 248


```

|||||.....
717 GGAACGCGAAGAGATCTAAGAGAGCTGCTCCCTGCTTATACCT 766
249 ThGluValThrLeuGluValGluProValGluMetLeuValGlu 265
767 GCGAAAAAGTGGGTGGAGTGAACCCGTTGGACTGTGAGGAAG 816
265 yAspArgValGluLeuArgCysLeuAlaAspGlyAsnProProHisP 282
817 GGACCATGTGAAGATCAGGTGTGACTGACGGCAACCTCAACCCACT 866
282 heSerIleSerLysGlnAsnProSerThrArgGluAlaGluGluThr 298
867 TCACATATCAACAGAAAGAACCCAGCATGAGAGATGAGAGGGAGAC 916
299 ThrAsnAspAsnGlyValLeuValLeuGluProAlaArgLysGluHis 315
917 ACCGATGAAAACGGTCTCTGCTGCTTGGAGCCGCCACAGACACAG 966
315 rGlyArgTyrGluCysGlnGlyLeuAspLeuAspThrMetIleSerLeu 332
967 TGGGCTCTACCATGTACAGAGCTGACCTGGAACCTACAGATTATGCTGT 1016
332 euSerGluProGlnGluLeuLeuValAsnTyrValSerAspValArgVal 348
1017 CGAGTAGACCCCTGAGCTGTGTGTAACATGTGTGATGTTCACAGTG 1066
349 SerProAlaAlaProGluArgGlnGluGlySerSerLeuThrLeuTyrC 365
1067 GATCCAAACGCCCTGATAGTCCAGAGAGTGAAGCTTACACTACCTG 1116
365 sGluAlaGluSerSerGlnAspLeuGluThrGlnTyrLeuArgGluT 382
1117 TTAGGCGAGAGTAAACAGACCTTGAGTTGAGTGGCTGAGACAGACA 1166
382 hGlyGlnValLeuGluArgGlyProValLeuGlnLeuHisAspLeuLys 398
1167 CAGGCCAGCTTCTGGAAAAGGGCCCATCTTCAAGTAAACAACGGAAA 1216
399 ArgGluAlaGlyGlyTyrArgCysValAlaSerValProSerIlePr 415
1217 AGGAGACAGGGGAGAGATATCTGTGTGATGTGCCAGAGGTTCC 1266
415 oGlyLeuAsnArgThrGlnLeuValAsnValAlaIlePheGlyProP 432
1267 TGGCTTGAATCCACACGCGGCTCAGCGTGGCAATTTTGGTCCCT 1316
432 rPheAlaPheLysGluArgLysValThrValLysGluAsnMetValLeu 448
1317 GGATGGCACCAAGAGAGAGAGTGTGGGCCAAGAGATCAATGCTG 1366
449 AsnLeuSerCysGluAlaSerGlyHisProArgProThrIleSerTrpAs 465
1367 AATGTGTCTTGAGGCTTCAGGCCATCTCAGCCACCATCTCTCGGAA 1416
465 nValAsnGlyThrAlaSerGluGlnAspGlnAspProGlnArgValLeu 482
1417 TATCAATGGCTCGCAACATGAATGAACCTCATCTCAGACAGATGGA 1466
482 eThrLeuAsnValLeuValThrProGlnLeuLeuGluThrGlyValGlu 498
1467 GCACCTGAAATGCTTGTGACCCAGAGCTTCTGAGACCCGGTCCAGAA 1516
499 CysThrAlaSerAsnAspLeuGlyLysAsnThrSerIleLeuPheLeuG 515
1517 TGTACAGCTTCAACCTCTGGGCTCATATCACACAGGTCATCATCTGAA 1566
515 uLeuValAsnLeuThrThrLeuThrProAspSerAsnThrThrGlyL 532
1567 GCTGTGCACTTTAACCACTTCACACATGACACAGCAAAACCACTGGCC 1616
532 euSerThrSerThrAlaSerProHisThrArgAlaAsnSerThrSerThr 548
|||||.....

```

```

1617 TCAGACCCCCACAGTCAGTCTCACAGACAGCAACAGACCTCCACA 1666
549 GluArgLysLeuProGluProGluSerArgGlyValValIleValAla 565
1667 GAGAAAAAGTCCACAGACAGAGCAAAAGTGTGGTCATCTGGCTGT 1716
565 lIleValCysIleLeuValLeuAlaValLeuGlyAlaValLeuTyrPhe 582
1717 GATACTATGTAGCGGTGGTGTGCTGTGCTGCTGAGGCTACCTCATACT 1766
582 euTyrLysGlyGlyLysLeuProCysArgArgSerGlyLysGlnGlu 598
1767 TCTACAAAGAGGCCAACTGCTTGTGGACGCTCAGGAAAAACAGAGAT 1815
599 ThrLeuProProSerArgLysSerGluLeuValValGluValLysSer 615
1815 .....
615 PlysLeuProGluGluMetGlyLeuLeuGlnGlySerSerGlyAspLys 632
1815 .....
632 rGAlaProGlyAspGlnGlyGlyTyrIleAspLeuArgHis 646
1816 .....GGAGAGAAATACATCATGTGAGGCAT 1842
seq_name: gb_htg:AP001557
seq_documentation block:
LOCUS AP001557 182429 bp DNA linear HTG 03-NOV-2000
DEFINITION Homo sapiens chromosome 11 clone RP11-680A7 map 11q23, WORKING
DRAFT SEQUENCE, 21 unordered pieces.
ACCESSION AP001557 GI:11094164
VERSION AP001557.3
KEYWORDS HTG, HTGS, PHASRL, HTGS, DRAFT.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 182429)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 182,429 genomic DNA of 11q23
Published Only in Database (2000) In press
2 (bases 1 to 182429)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submision
Submitted (29-MAR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@psc.riken.go.jp, URL:http://hpg.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
On Nov 3, 2000 this sequence version replaced gi:8117391.
COMMENT
----- Genome Center
Center: RIKEN Genomic Sciences Center (GSC)
Center code: RIKEN
Web site: http://hpg.gsc.riken.go.jp/
Contact: hattori@psc.riken.go.jp
----- Project Information
Center project name: HumDraT11
Center clone name: RP11-680A7
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: dye-terminator ET-chemistry; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 175597 bases at least Q40
Consensus quality: 178555 bases at least Q30
Consensus quality: 179768 bases at least Q20
Insert size: 180429; sum-of-ctrls
Quality coverage: 9.01x in Q20 bases; sum-of-ctrls

```

NOTE: This is a 'working draft' sequence. It currently consists of

21 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```
1 30107 contig of 30107 bp in length
30208 54357 contig of 24150 bp in length
54458 69839 contig of 15382 bp in length
69940 84048 contig of 14109 bp in length
84149 95177 contig of 11029 bp in length
95278 10763 contig of 12486 bp in length
107864 119057 contig of 11194 bp in length
119158 128134 contig of 8697 bp in length
128235 136903 contig of 8884 bp in length
137004 145887 contig of 7212 bp in length
145988 153199 contig of 6266 bp in length
153300 159565 contig of 3915 bp in length
159666 163580 contig of 1472 bp in length
163681 165152 contig of 1472 bp in length
165153 165252 contig of 100 bp in length
165253 166901 contig of 1649 bp in length
166902 167001 contig of 100 bp in length
167002 170460 contig of 3459 bp in length
170461 170560 contig of 100 bp in length
170561 172761 contig of 2201 bp in length
172762 172861 contig of 100 bp in length
172862 175605 contig of 2744 bp in length
175606 175705 contig of 100 bp in length
175706 178308 contig of 2603 bp in length
178309 178408 contig of 100 bp in length
```

NOTE: This is a 'working draft' sequence. It currently consists of 21 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```
1 30107 contig of 30107 bp in length
30108 30207 contig of 100 bp in length
30208 54357 contig of 24150 bp in length
54358 54457 contig of 100 bp in length
54458 69839 contig of 15382 bp in length
69840 69939 contig of 100 bp in length
69940 84048 contig of 14109 bp in length
84049 84148 contig of 100 bp in length
84149 95177 contig of 11029 bp in length
95178 95277 contig of 12486 bp in length
95278 107763 contig of 100 bp in length
107764 107863 contig of 11194 bp in length
107864 119057 contig of 11194 bp in length
119058 119157 contig of 8697 bp in length
119158 128134 contig of 8697 bp in length
128135 128234 contig of 100 bp in length
128235 136903 contig of 8669 bp in length
136904 137003 contig of 100 bp in length
137004 145887 contig of 8884 bp in length
145888 145987 contig of 100 bp in length
145988 153199 contig of 7212 bp in length
153200 153299 contig of 100 bp in length
153300 159565 contig of 6266 bp in length
159566 159665 contig of 100 bp in length
159666 163580 contig of 3915 bp in length
163581 163680 contig of 100 bp in length
163681 165152 contig of 1472 bp in length
165153 165252 contig of 100 bp in length
165253 166901 contig of 1649 bp in length
166902 167001 contig of 100 bp in length
167002 170460 contig of 3459 bp in length
170461 170560 contig of 100 bp in length
170561 172761 contig of 2201 bp in length
172762 172861 contig of 100 bp in length
172862 175605 contig of 2744 bp in length
175606 175705 contig of 100 bp in length
175706 178308 contig of 2603 bp in length
178309 178408 contig of 100 bp in length
```

FEATURES
source 178409 180918: contig of 2510 bp in length
* 180919 181018: gap of 100 bp
* 181019 182429: contig of 1411 bp in length.
Location/Qualifiers
1. 182429
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q23"
/clone="RP11-680A7"
1. 30107
/note="assembly-fragment"

```
misc_feature 30208..54357  
/note="assembly-fragment"  
misc_feature 54458..69839  
/note="assembly-fragment"  
misc_feature 69940..84048  
/note="assembly-fragment"  
misc_feature 84149..95177  
/note="assembly-fragment"  
misc_feature 95278..107763  
/note="assembly-fragment"  
misc_feature 107864..119057  
/note="assembly-fragment"  
misc_feature 119158..128134  
/note="assembly-fragment"  
misc_feature 128235..136903  
/note="assembly-fragment"  
misc_feature 137004..145887  
/note="assembly-fragment"  
misc_feature 145988..153199  
/note="assembly-fragment"  
misc_feature 153300..159565  
/note="assembly-fragment"  
misc_feature 159666..163580  
/note="assembly-fragment"  
misc_feature 163681..165152  
/note="assembly-fragment clone_end:T7 vector_side:left"  
misc_feature 165253..166901  
/note="assembly-fragment clone_end:SP6 vector_side:left"  
misc_feature 167002..170460  
/note="assembly-fragment"  
misc_feature 170561..172761  
/note="assembly-fragment"  
misc_feature 172862..175605  
/note="assembly-fragment"  
misc_feature 175706..178308  
/note="assembly-fragment"  
misc_feature 178409..180918  
/note="assembly-fragment"  
misc_feature 181019..182429  
/note="assembly-fragment"
```

BASE COUNT 45724 a 42083 c 43875 g 48747 t 2000 others
ORIGIN

alignment_scores: Length: 870
Ratio: 4.264 Gaps: 15
Percent Similarity: 51.379 Percent Identity: 51.264

alignment_block:
US-09-653-961-2 x AP001557/rev ..

Align seg 1/1 to reverse of: AP001557 from: 1 to: 182429

```
188 ValHisIleGlnSerGlnThrValGluSerSerGlyLeuTyrThr 204  
|||||  
29642 GTCCACATTCAGTGTCCACAGCTGTGAGTTCAGTGTTCACCTT 29593  
|||||  
204 uGlnSerIleLeuTyrSalGlnLeuValGlyGluAspLysAspAlaGlnP 221  
|||||  
29592 GCAGAGTATTCGAGGACAGCTGTGAAGAGCAAGCAAGATGCCAGT 29543
```

```

221 hefYrCysGluLeuAanTyrArgLeuProSerGlyAsnHisMetLysGlu 237
222 |||||||
223 29542 TTACTGTGAGCTCACTACCGGCTCCCGTGGAGAACCATGAAAGAG 29493
224 |||||||
225 238 SerArgGluValThrValProValPhe..Tyr..... 247
226 |||||||
227 29492 TCCAGGGAAGTCACCGTCCCTGTTTCTGTGTAGTACTAACCTGACTCTCT 29443
228 |||||||
229 247 ..... 247
230 29442 GCTTACTGGGGCTGGGCTGCTGGAAGTCTGCTGCTGCCCTTATCC 29393
231 |||||||
232 247 ..... 247
233 29392 AGCCGAGGTTGACTCATGGGCAAGAGCTACCCCTTTGCTATGCAGA 29343
234 |||||||
235 248 .ProThrGluLysValTrpLeuGluValGluProValGlyMetLeuLysG 264
236 |||||||
237 29342 CCCGACAGAAAAAGTGTGGCTGGAAGTGAAGCCCGTGGGAATGCTGAAG 29293
238 |||||||
239 264 LuGlyAspArgValGluIleArgCysLeuAlaAspGlyAsnProProPro 280
240 |||||||
241 29292 AAGGGGACCGCGTGAATAATCAGGTGTTGCTGATGSCAACCTCCACCA 29243
242 |||||||
243 281 HisPheSerIleSerLys..... 286
244 |||||||
245 29242 CACTTCAGCATACACCAAGCAGGTGTGGGGCTTTGTGCTGGTGCAGGTG 29193
246 ..... 286
247 29192 GGGATGCAGACAGAGACCCAGGTGAGTGGCGGAGGTGCACCTCTCTCA 29143
248 .....Gln..AsnProSerThrArgGluAlaGluGluLut 298
249 |||||||
250 29142 CTTCCCAATCTCTCCCTAGAACCCACGACCAAGGAGGAGGAGAGAGA 29093
251 |||||||
252 298 hrThrAsnAspAsnGlyValLeuValLeuGluProAlaArgLysGluHis 314
253 |||||||
254 29092 CAACCAACGCAACAGGGGTCTGTGTCTGTGAGCTGCCCGGAGAGAAAC 29043
255 |||||||
256 315 SerGlyArgTyrGluCysGlnGlyLeuAspLeuAspThrMetIleSerLe 331
257 |||||||
258 29042 AGTGGGCGCTATGATATGTCAGGGCGCTGAGCTTGACACCATGATATCGCT 28993
259 |||||||
260 331 uLeuSerGluProGlnGluLeuLeuValAsn..Tyr..... 342
261 |||||||
262 28992 GCTGAGTGAACCAAGAACTACTGTGACTGTGAGGGGCTGGGGAGCC 28943
263 ..... 342
264 28942 AGGACAGGGGACACAGCTGGGGCAATAGGGAAGTGGCCCACTGCC 28893
265 .....ValSerAspValArgValSerP 350
266 |||||||
267 28892 TGGGTGACATGATGCTGTACCCAGATGTCTCTGACGTCCGAGTGTGTC 28843
268 |||||||
269 350 roAla..AlaProGluArgGlnGluGlySerSerLeuThrLeuThrCysG 366
270 |||||||
271 28842 CCGGAGGCGCTGTGAGAGCAAGAAAGCAGAGGCTCACCTGACCTGTGTA 28793
272 |||||||
273 366 uAlaGluSerSerGlnAspLeuGluPheGlnTrpLeuArgGlu..... 381
274 |||||||
275 28792 GGCAGAGAGTGAAGCAAGACCTCGAGTTCAGTGGCTGAGAGAGAGGTAC 28743
276 ..... 381
277 28742 CCAAGCGGCTGTGGGCAAGGGCTGTAGGGTGGGGCTGGCGGTG 28693
278 .....Th 382
279 382 .....Th 382
280 28692 ACAAGATGTGTCTTTCCAGACCTCTGACCATCTCTATCCAGAC 28643

```

```

382 rGlyGlnValLeuGluArgGlyProValLeuGlnLeuHisAspLeuLysA 399
383 |||||||
384 28642 AGGCCAGGTGTGGAAGAGGGGCTGTGCTTCAAGTTCATGACCTGAAC 28593
385 |||||||
386 399 rGlyAlaGlyGlyGlyTyrArgCysValAlaSerValProSerIlePro 415
387 |||||||
388 28592 GGGAGCAGAGAGCGGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 28543
389 |||||||
390 416 GlyLeuAsnArgThrGlnLeuValAsnValAlaIlePhe..... 428
391 |||||||
392 28542 GGCCTGAACCGACACAGCTGTCAACGTGSCATTTTGTGTAGGGCT 28493
393 ..... 428
394 28492 CCTCTGGGTAGAGACCAAGCTACCCCAAGTGGGGTGTTTTAAGCTCT 28443
395 ..... 428
396 428 ..... 428
397 28442 TTGACAAAAAGCCACCTGCTGCTGGGAGCTGTGTGGGAGGGGAG 28393
398 .....GlyPr 430
399 28392 GCAGGCTAGCTGGGCTTGCCACCTCACTCTCTCTCTCTCTCTCTCTCT 28343
400 |||||||
401 430 oProThrMetAlaPheLysGluArgLysValTrpValLysGluAsnMetV 447
402 |||||||
403 28342 CCTTGTGATGGCATTCAGAGAGAGAGAGGTGTGGTGAAGAGCATATG 28293
404 |||||||
405 447 AlLeuAsnLeuSerCysGluAlaSerGlyHisProArgProThrIleSer 463
406 |||||||
407 28292 TGTGATGTGTCTGTGTGAAGGCTCAGGGCACCCCGGCCACATCTCC 28243
408 |||||||
409 464 TrpAsnValAsnGlyThr..... 469
410 |||||||
411 28242 TGAAGCTCAACGCAAGGCTGAGGGGTGTCTCATCTGCACACTCTCTCT 28193
412 ..... 469
413 469 ..... 469
414 28192 AATCCTGGCTGGCTTTTGTGGCCAGACAGAGATCTCCCATCCCT 28143
415 .....AlaSerGluGlnAspGlnAspPro 477
416 |||||||
417 28142 GAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 28093
418 |||||||
419 478 GlnArgValLeuSerThrLeuAsnValLeuValThrProGluLeuG 494
420 |||||||
421 28092 CAGCGAGTCTGAGCACTTGAATGTCTCTGTACCCCGGAGCTGTGGA 28043
422 |||||||
423 494 uThrGlyValGluCysThrAlaSerAsnAspLeuGlyLysAsnThrSer 511
424 |||||||
425 28042 GACAGGTGTATGATGACAGGGCTCCAGAGACTGGGCAAAAACACCA 27993
426 |||||||
427 511 LeLeuPheLeuGluLeu..Val..... 517
428 |||||||
429 27992 TCCTTCTCTGAGCTGGGTGAGGGCTGCATCCCTGAGGGATGACAG 27943
430 ..... 517
431 27942 GGCAGGCCAGGACGGGCAAGTACTTGCCCTCTCTCTCAACCCCTCC 27893
432 .....AsnLeuThrThrLeuThrProAspSerAsnThrThr 529
433 |||||||
434 27892 TCTGCTCTCTGATGATTAACCACTTCAACAGACCTCAACCAAC 27843
435 |||||||
436 530 ThrGlyLeuSerThrSerThrAlaSerProHisThrArgAlaAsnSer 546
437 |||||||
438 27842 ACTGGCTCTGACACTTCCACTGCACTCTCTACATACAGCCACAGCAG 27793
439 ..... 548
440 546 rSerThr..... 548
441 |||||||
442 27792 CTCACAGGTAAGCCAGGCTTGCAAGAGAACAGGGCTGTGCCAGGCA 27743
443 ..... 548
444 548 ..... 548

```


[illegible]

```
248 rothGluValTrpLeuGluValGluProValGlyMetLeuLysGlu 264
|||||
33157 CGACGAAAAAGTGTGGCTGGAGTGGAGCCCTGGGGAATGCTGAAGAA 33108
265 GlyAspArgValGluIleArgCysLeuAlaSpGlyAsnProProH 281
|||||
33107 GGGACCGCGTGGGAATCAGGTGTGGCTGATGGCAACCTCCACCACA 33058
281 sPheSerIleSerLys..... 286
|||||
33057 CTTGAGCATCAGCAAGCAGGTGTGGGGCTTTGCTGGTGACAGTGGG 33008
286 ..... 286
33007 GATGCACAGACAGACCAGGTGGAGTGGGGAGTGCATCTGCTACC 32958
287 .....Gln..AsnProSerThrArgGluAlaGluGluThr 298
|||||
32957 TCCCAATCCTCCTAGAACCCCGACACCAGGAGCAGAGAGAGACA 32908
299 ThrAsnAspAsnGlyValLeuValLeuGluProAlaArgLysGluHisP 315
|||||
32907 ACCAAGCAACAGGGGCTCCTGGTCTGGAGCTGCCGGAAGAACACAG 32858
315 rGlyArgGlyArgLysGlnGlyLeuAspLeuAspThrMetIleSerLeuL 332
|||||
32857 TGGGGCTATGATGTACAGGGCTTGACTGGACCATGATATGCTGC 32808
332 eusSerGluProGlnLeuLeuValAsn..... 341
|||||
32807 TGAGTGAACCAAGGAACACTGCTGTAACCTGTGAGGGGCTGGGACCAG 32758
341 ..... 341
32757 GACAGGGGACAGAGCTGGGCAATAGGGAACCTGCCCTGCCCTG 32708
342 .....Tyr..ValSerAspValArgValSerPr 350
|||||
32707 GGTGACATGATGTCTGTACCCCAAGATGTGTGACGTCCGATGATGC 32658
350 oAlaAlaProGluArgGlnGlnGlySerSerLeuThrLeuThrCysGlu 366
|||||
32657 CGCAGCCCTGAGAGAGACAGGAAGGCGACGCTCACCTGACCT..GTGA 32611
367 AlaGluSerSerGlnAspLeuGluPheGlnTrpLeuArgGlu..Glu.... 381
|||||
32610 GCAGAGAGTAGCCAGAGGCTCGAGTTCAGAGGCTGAGAGAAAGGTACC 32561
381 ..... 381
32560 CAGCGGGGTCTGGGGCAAGGGGCTGATGGGTGGGTGGCGGGTG 32511
382 .....Th 382
32510 ACAAGATGTGCTTTCCAGACCTGTGACCCATCTCATCCAGAC 32461
382 rGlyGluValLeuGluValArgGlyProValLeuGlnLeuHisAspLeuLys 399
|||||
32460 AGGCCAGGTGCTGGAAGGGGGCTGTGCTTCAGTTGCATGACCTGAAC 32411
399 rGluAlaGlyGlyGlyTyrArgCysValAlaSerValProSerIlePro 415
|||||
32410 GGGAGGAGAGAGGGGCTATCGCTGCTGGCTGTGCCCCACATACCC 32361
416 GlyLeuAsnArgThrGlnLeuValAsnValAlaIlePheGly..... 429
|||||
32360 GGGCTGAACCGCACAGCTGTGTAACGTGCCATTTTGTGTAGGCCCT 32311
429 ..... 429
32310 CCTCTGGGTAGAGACCAAGTCAACCCCAAGTGGGTGTTTAACTC 32261
429 ..... 429

429 ..... 429
32260 TTTGACAAAAAGCCACTGTGCTGGGAGCTCTGTGGCAGGGGA 32211
430 .....Pr 430
32210 GCGAGGCTAGCTGGGCTTGCCACCTCACCCTTCCTACCTCCCAAGGCC 32161
430 o.ProThrMetAlaPheLysGluArgLysValTrpValLysGluAsnMet 446
|||||
32160 CCGCTGGATGGCATTCACAGAGAGAGAGAGGTGGGTGAAGAAGATATG 32111
447 ValLeuAsnLeuSerCysGluAlaSerGlyHisProArgProThrIleSe 463
|||||
32110 GTGTGAATCTGCTGTGTGAAGGCTCAGGGACCCCGGCCACCATCTTC 32061
463 rTrpAsnValAsnGlyThr..... 469
|||||
32060 CTGGAACGTACAGGCGACGGTGAAGCGGGTGTCTCATCTGCCACATCTCC 32011
469 ..... 469
32010 TAATCCCTGGCTGGGCTTTGTGGCCAGAGACAGAGATCTCCCATCCCC 31961
470 .....AlaSerGluGlnAspGlnAspPr 477
|||||
31960 TGAGCTGCTCTCTCTCTTCCTGCGACAGGCAAGTGAACAAAGCAAGATCC 31911
477 oGlnArgValLeuSerThrLeuAsnValLeuValAlaThrProGlnLeuLeu 494
|||||
31910 ACACGAGTCTCTGACACCTGAAATGTCTCTGTGACCCCGGAGCTGTGG 31861
494 IuThrGlyValGluCysThrAlaSerAsnAspLeuGlyLysAsnThrSer 510
|||||
31860 AGACAGGTGTGAATGCACGGGCTCCAAAGACCGTGGCAAAAACACAGC 31811
511 IleLeuPheLeuGluLeu..Val..... 517
|||||
31810 ATCCCTCTTCGTGAGCTGGGTGAGGGCTGCATCCCTCAGAGGATGGCAG 31761
517 ..... 517
31760 GGGCAGGCCGACGAGCGGGCAGTAACTTGGCTTCTCTTAACCCCTCC 31711
518 .....AsnLeuThrThrLeuThrProAspSerAsnThrTh 529
|||||
31710 CTCTGCTCTCTAATCAATTTAACACCCCTCACACAGACTCCAAACACAC 31661
529 rThrGlyLeuSerThrSerThrAlaSerProHisThrArgAlaAsnSerT 546
|||||
31660 CACTGGCTCAGCACTTCCACTGCCAGTCTCATACAGAGACCAACAGCA 31611
546 hrSerThr..... 548
|||||
31610 CCTCCACAGGTAAAGCAGGCTGGCAAGAGAAACAGGGCTGTGCCAGGCA 31561
548 ..... 548
31560 TCCCTTTTGCCTGTTCCTCCCCAGAGAGCCCTGTCCAGAAAGGTGAGTA 31511
548 ..... 548
31510 GCAGCCCATCTTTGCGGGCTGGAAGTGGGGCAACGATGTGAGCA 31461
549 .....GluArg 550
|||||
31460 AGTGGCTTGGGCGAGGAGTGAAGAGAGGTCTTTTGTGGCGCAGAGAGA 31411
551 LysLeuProGluProGluSerArg..GlyValAlaIleValAlaValIleV 567
|||||
31410 AMGCTCCGAGAGCGGAGAGCGGGGGGCTGTGATCTGATGATGATG 31361
567 aLysIleLeuValLeuAlaValLeuGlyAlaValLeuThrPheLeuTyr 583
```


248 ProThrGluLysValTrpLeuGluValGluProValGlyMetLeuLysGlu 264
 450 CCGACGAGAAAAGTGTGCTGGAGTGGAGCCCGTGGGAATGCTGAAGA 499
 264 uGlyAspArgValGluIleArgCysLeuAlaAspGlyAsnProProH 281
 500 AGGGGACCGCGTGAATCAGGTGTGGCTGATGGCAACCTCCACAC 549
 281 AspSerIleSerLys..... 286
 550 ACTTCAGCAATCAGACAGAGAGTGTGGGGCTTGTGCTGGTGAAGTGG 599
 286 286
 600 GATGCAAGACAGACAGCCAGGTGAGTGGCGGAGGTGCATCTGCTCAC 649
 287Gln..AsnProSerThrArgGluAlaGluGluLuth 298
 650 CTCGCCAATCTCTCCCTAGAACCCACAGCAGGAGGAGAGAGAGAGAC 699
 298 rThrAsnAspAsnGlyValLeuValLeuGluProAlaArgLysGluHis 315
 700 AACCAAGACAGACAGGGGTCTGGTGGAGCTGCCCGGAGAGACACA 749
 315 eArgLysArgLysGluGlnGlyLeuAspLeuAspThrMetIleSerLeu 331
 750 GTGGGCGCTATGATGATGAGCGCTGGAACTGGACACCATGATATGCTG 799
 332 LeuSerGluProGlnGluLeuLeuValAsn..... 341
 800 CTGAGTACACACAGAGAACTAGTGTGTAAGTGTGAGGGGCTGGGACCA 849
 341 341
 850 GGACAGGGGAGACAGCGTGGGGCAATAGGGAAGTGGCCCACTGCTGCT 899
 342Tyr..ValSerAspValArgValSerP 350
 900 GGGTACATGATGTGCTGATACCCCAAGATGTCTGAGCTCGAGTGAATC 949
 350 roAlaAlaProGluArgGlnGluLysSerLeuThrLeuThrCysGlu 366
 950 CCGCAGCCCTGAGAGACAGAGAGGACAGCCCTACCTGAGTGTGAG 999
 367 AlaGluSerSerGlnAspLeuGluPheGlnTrpLeuArgGluLys 381
 1000 GCAGAGATAGCCAGGACCTGAGTTCAGTGGCTGAGAGAGAGATACC 1049
 381 381
 1050 CAGGGGGGTGTGGGCAAGGGGCTGATGGGTGGGTGGGCTGGCGTGA 1099
 382Thr 382
 1100 CAGAATGTGCTCTTCCAGACCTGTGACCATCTCATCCAGACA 1149
 383 GlyGlnValLeuGluArgGlyProValLeuGlnLeuHisAspLeuLysAr 399
 1150 GACCAAGTCTCGAAGAGGGGCTGTGCTGATGATGACATGAAACG 1199
 399 gGluAlaGlyGlyLysArgCysValAlaSerValProSerIleProG 416
 1200 GGAGGAGAGAGCGGCTATCGCTGGTGGCTGTGCTGCCACATATCCG 1249
 416 LysLeuAsnArgThrGlnLeuValAsnValAlaIle..Phe..... 428
 1250 GCCTGAACCGCACACAGCTGTGCAAGCTGCCATTTTGTGATAGGCCCTC 1299
 428 428
 1300 CCTGTGGTATAGACACAGGTCAACCCCAAGTGGGTGTTTAAAGCTCTT 1349
 428 428

1350 TGACAAAAGCCACCTGCTGCGCTGGGAGCTGTGTCGGAGGGGAGG 1399
 429GlyProP 431
 1400 CAGGCTAGCTGGGCTTGACCTACCTCTCCCTACCTCCCAAGGGCCCC 1449
 431 roTPMetAlaPheLysGluArgLysValTrpValLysGluAsnMetVal 1447
 1450 CTGGATGGCATTCAGAGAGAGAGAGGTGTGGTGAAGAAGAAATATGCTG 1499
 448 LeuAsnLeuSerCysGluAlaSerGlyHisProArgProThrIleSerTr 464
 1500 TTGAATCTGCTGTGAGAGCGTCAGAGGACCCCGCGCCACCATCTCTG 1549
 464 PAsnValAsnGlyThr..... 469
 1550 GAACGTCAACGGCACGGGTGAGCGGGTCTCTATCTGCCACACTCTCTAA 1599
 469 469
 1600 TCCCTGCTGGCGCTTGTGGCAGAGACAGATCTCCCATCCCTGA 1649
 470AlaSerGluGlnAspGlnAspPr 477
 1650 GCTGCTCTCTCTCTCTCTCTCTCTCGCAGCAAGTGAACAGACAGATCC 1699
 477 oGlnArgValLeuSerThrLeuAsnValLeuValThrProGluLeuG 494
 1700 ACAGCAGTCTGAGCAGCAGCAGTCAATGCTGTCGACCCGGAGGTGTGG 1749
 494 LuThrGlyValGluCysThrAlaSerAsnAspLeuGlyLysAsnThrSer 510
 1750 AGCAGGTGTGAATGCACGGCTCCAGACCTGGGCAAAAAACACAGC 1799
 511 IleLeuPheLeuGluLeu..Val..... 517
 1800 ATCTCTCTCTGAGCTGGGTGAGGCTGCATCCCTGCAGGGCATGGGCA 1849
 517 517
 1850 GGGGCTCAGGGCCAGGACAGGGCAGTAACTGCCGCTTCTTAACACCC 1899
 518AsnLeuThrThrLeuThrProAspSerAsnTh 528
 1900 TCCCTGCTCTCTCTCTCTCTCTCTCTGATTAACCACTTACACAGACTCCACAC 1949
 528 rThrThrGlyLeuSerThrSerThrAlaSerProHisThrArgAlaAsn 545
 1950 AACCACTGGCTCAGCAGCTTCCAGTCCAGTCTCATACAGAGCCACCA 1999
 545 eThrSerThr 548
 2000 GCACCTCCACA 2010

seq.name: gb.ro:MMUC18GP
 seq.documentatation_block: 6102 bp DNA linear ROD 05-OCT-1993
 LOCUS MMUC18GP
 DEFINITION M.musculus gene for MUC18 glycoprotein (partial).
 ACCESSION X74628
 VERSION X74628.1 GI:406585
 KEYWORDS immunoglobulin gene superfamily; melanoma associated protein; MUC18 glycoprotein.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 6102)
 AUTHORS Rothbacher,U.
 DIRECT SUBMISSION
 TITLE Submitted (17-AUG-1993) U. Rothbacher, Institute for Immunology,
 JOURNAL University of Muenich, Goethestr 31, 80336 Muenchen 2, FRG

REFERENCE	AUTHORS	TITLE
2 (bases 1 to 6102)	Rothbaeher, U., Sers, C., Rietmuller, G. and Johnson, J. P.	Characterization of the human melanoma metastasis associated

1526 CTCATCTGGGCAAGAGGNACTGGCTCGTCTGTCAGGCTGACATTTT 1575
343ValSerAspValArgValSerProAlaIaIaPr 353
1576 TTTCGTATGCTCCAGATGTCTGATGTTCAAGTGAATCCAACTGCCCC 1625
353 oGluArgGlnGluGlySerSerLeuThrLeuThrCysGluAlaGluSer 370
1626 TGAAGTCCAGGAGGTGAGAGCTCAGCTGACTGCGAGGAGAAAGTA 1675
370 eGlnAspLeuGluPheGlnTrpLeuArgGlu.Glu..... 381
1676 ACCAGAGACTTGAGTTGAGTGCTGAGAGCAAGTAACCAACACACTT 1725
381 381
1726 GAGACTGGGGAGAGGGGACATGAGACTGCTCTTTTCAGACCCCT 1775
382ThrGlnValLeuGluArgGlyProVal 391
1776 GACCCACTCTCACCCCTAGACAGGCCAGCTGCGGAAGGTCCTCCGTC 1825
392 LeuGlnLeuHisAspLeuLysArgGluAlaGlyGlyTyArgCysVa 408
1826 CTCACGCTAAACACGCTGACGAGCGGAGACGAGGCGGTATCTGCA 1875
408 AlaSerValProSerLeuProGlyLeuAsnArgThrGlnLeuValAsn 425
1876 GGCATGTCGCCAGAGTTCGTGCTTGATCCACGCTGGTCAGCG 1925
425 AlaAlaIlePheGlyPro..... 430
1926 TGGGCATTTTGGTGAGCCCTTTGACAGTACACGAGTCAACCTGACTT 1975
430 430
1976 GTGTGTTTGAAGCTGTGGCAGCTTGTACTGTGGAGCTGAGCGGGA 2025
431P 431
2026 GCAGGCTATTGGTATTGGGGGCTAAACCTCCCTGCCCAAGGCTC 2075
431 ro..TriPheAlaPheLysGluArgLysValTrpValLysGluAsnMet 447
2076 CCCATGATGCAATTAAAGAGAGAGAGGTGGGTGCAAGAAATGCG 2125
447 alleuAsnLeuSerCysGluAlaSerGlyHisProArgProThrIleSer 463
2126 TGCTGAATCTGTCTTGTGAGGCTTCAGGACATCTCAGCCCACTCTCC 2175
464 TrpAsnValAsnGly..... 468
2176 TGAATGTCATAGTGTGCTTCGGTTCGAGCCGGCTTGCTTCATCTCCCAT 2225
468 468
2226 TCTCTAATCCCTCGTAGCCTTGAGCCAGGCCAGAGATCTCCGCC 2275
469Thr..AlaSerGluGlnAspProG 478
2276 CTCTGAGAGGCCCATTTACTTCTAGGCACTGAATGGAATCCAGATCCAC 2325
478 InArgValLeuSerThrLeuAsnValLeuValThrProGluLeuGlu 494
2326 AGACAGTGTGACACCTTGAATGCTTGTGAGCCAGAGACTTCTGGAG 2375
495 ThrGlyValGluCysThrAlaSerAsnAspLeuLysAsnThrSerI 511
2376 ACAGGTGACAGGTGACAGCTCCAACTCCCTGGGCTCAAAACACACAC 2425
511 e..... 511
2426 CATTTGTTGAAGCTGGGTGAGAGCGGACCTCTGACAGGCTGGCAGG 2475

512Leuphe 513
2476 GCATGCCAGGACAGGCAATTAATCTGTGCTACTAAACCCCAACTCCT 2525
514 LeuGlu..LeuValAsnLeuThrThrLeuThrProAspSerAsnThrTh 530
2526 CTGCTCTTAATCACTTAACCAACCCCATACCTGACTCAGCCAAACCA 2575
530 hrGlyLeuSerThrSerThrAlaSerProHisThrArgAlaAsnSerThr 546
2576 CTGGCTCAGCACCCCTCACAGTCACTCTCAACACAGGCAACAGACACC 2625
547 .Ser..... 547
2626 TCACAGGTAAAGCGGCGCTCGCAGATGTAACCACTCACTCCCTTT 2675
547 547
2676 TGAGAAAGTTAAGCGCCCATCTGTCCCTGGGCTGAGAGGTTGGGCA 2725
547 547
2726 GCAACGCAATTTGCCAGCCAGGAGGAGGAATGACTACAAGTCTTTG 2775
548ThrGluArgLysLeuProGluProGluSerArgGlyValAlley 563
2776 TTTCACAGAGAAAGAGCTGCCACAGCCAGAGCAAGAGTGTGCTATCG 2825
563 AlaValAlleyAlaLysIleLeuValLeuAlaValAlleyValLeu 579
2826 TGCTGTGATGATGTGTACTGTGTCTGCTCTGCTGCTGCTGCTGCTC 2875
580 TyrPheLeuThrLysLysGlyLysLeuProCysArgArgSerGlyLys 596
2876 TATTCTTCTACAGAGGCAAGCTGCATGTGAGCTCGGGAACA 2925
596 nGlu..Ile..... 598
2926 GGAGATGTAAGCCAGGCTTATGCTCTCCCTCCACCCCACTTTTG 2975
598 598
2976 GCGTGCCAGCTGCTGTCTTAAGGCGCTGCTACCACTGCCAT 3025
598 598
3026 ACCGTGAGAGAGAAACTGAGCTGTTTATTTATGTCACACCAACTG 3075
598 598
3076 CCACCAACACACCACTTCCCAATTCCTCTCTGAGCGGCCCA 3125
598 598
3126 AAAGCAGAGTATCCATTCTTAAGAGTCCCTCCCTCCCAAGCTCC 3175
599ThrLeup 601
3176 CCCAGCCCCGAGTTGCCAACTCTCACTCCCTTCCAGAGAGCTTC 3225
601 roProSerArgLysSerGlnLeuValAlaGluValLysSerAspLysLeu 617
3226 CCCGAGCTGTAAGAGTAATTTGTAGTGAAGTAAAGTAAAGCTC 3275
618 ProGluGluMetGlyLeuLeuGlnGlySerSerGlyAspLysArgAlaPr 634
3276 CCAGAAAGATGCTCTCTTCAAGGAGCAACGAGTGAAGAGGCTCC 3325
634 oGlyAspGln 637
3326 AGGAGACCAAG 3335

seq_name: gb_hcg:AF002888

seq_documentation_block: 156869 bp DNA linear HTG 18-OCT-2000

LOCUS AP002888

DEFINITION Homo sapiens chromosome 11 clone RP11-122H4 map 11q, WORKING DRAFT

SEQUENCE, 25 unordered pieces.

ACCESSION AP002888

VERSION AP002888.1 GI:10880444

KEYWORDS HTG; HTGS-PHASE1; HTGS-DRAFT.

SOURCE Homo sapiens DNA, clone:RP11-122H4.

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 156869)

AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Matanabe,H. and Sakaki,T.,

TITLE Published Only in Database (2000) In press

JOURNAL 2 (bases 1 to 156869)

Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Matanabe,H. and Sakaki,T., Direct Submission

Submitted (16-OCT-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-15-1 Kitsato, Sagamihara, Kanagawa 228-8555, Japan Tel:81-42-778-9923, Fax:81-42-778-9924)

----- Genome Center

Center: RIKEN Genomic Sciences Center(GSC)

Center code: RIKEN

Web site: <http://hgp.gsc.riken.go.jp/>

Contact: hattori@hgp.gsc.riken.go.jp

----- Project Information

Center project name: RP11-122H4

Center clone name: RP11-122H4

----- Summary Statistics

Sequencing vector: PCR products: 100% of reads

Library: Dye-terminator ET-amersham; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 137843 bases at least Q40

Consensus quality: 147013 bases at least Q30

Insert size: 154469; sum-of-contigs

Quality coverage: 4.53x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 25 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs 'N', but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

1 16262 contig of 16262 bp in length

16363 32533 contig of 16171 bp in length

32634 47033 contig of 14400 bp in length

47134 60266 contig of 13133 bp in length

60367 70550 contig of 10184 bp in length

70651 83136 contig of 12486 bp in length

83237 90994 contig of 7758 bp in length

91095 99344 contig of 8250 bp in length

99445 106022 contig of 6578 bp in length

106123 112604 contig of 6482 bp in length

112705 123495 contig of 5570 bp in length

123596 130199 contig of 5121 bp in length

130200 130299 contig of 6604 bp in length

130300 132629 contig of 2230 bp in length

132630 132728 contig of 100 bp in length

132730 135477 contig of 2748 bp in length

135478 135577 contig of 100 bp in length

135578 139010 contig of 3433 bp in length

139011 139110 contig of 100 bp in length

139111 141568 contig of 2458 bp in length

141569 141668 contig of 100 bp in length

142232 142131 contig of 463 bp in length

142132 142231 contig of 100 bp in length

142232 144866 contig of 2635 bp in length

144867 144966 contig of 100 bp in length

144967 148414 contig of 3448 bp in length

148415 148514 contig of 100 bp in length

148515 150504 contig of 1990 bp in length

150505 150604 contig of 100 bp in length

150605 151939 contig of 1355 bp in length

151960 152059 contig of 100 bp in length

152060 153455 contig of 1396 bp in length

153456 153555 contig of 100 bp in length

153556 155251 contig of 1666 bp in length

155252 155351 contig of 100 bp in length

155352 155869 contig of 1518 bp in length.

Location/Qualifiers

1. 156869

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="11"

/map="11q"

/clone="RP11-122H4"

1. 16262

/note="assembly-fragment"

16363. 32533

/note="assembly-fragment"

32634. 47033

150605 151959 contig of 1355 bp in length

152060 153455 contig of 1396 bp in length

153556 155251 contig of 1666 bp in length

155352 156869 contig of 1518 bp in length.

NOTE: This is a 'working draft' sequence. It currently consists of 25 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of 'N', but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 16262: contig of 16262 bp in length

16263 16362: gap of 100 bp

16363 32533: contig of 16171 bp in length

32534 32633: gap of 100 bp

32634 47033: contig of 14400 bp in length

47034 47133: gap of 100 bp

47134 60266: contig of 13133 bp in length

60267 60366: gap of 100 bp

60367 70550: contig of 10184 bp in length

70551 70650: gap of 100 bp

70651 83136: contig of 12486 bp in length

83137 83236: gap of 100 bp

83237 90994: contig of 7758 bp in length

90995 91094: gap of 100 bp

91095 99344: contig of 8250 bp in length

99345 99444: gap of 100 bp

99445 106022: contig of 6578 bp in length

106023 106122: gap of 100 bp

106123 112604: contig of 6482 bp in length

112605 112704: gap of 100 bp

112705 118274: contig of 5570 bp in length

118275 118374: gap of 100 bp

118375 123495: contig of 5121 bp in length

123496 123595: gap of 100 bp

123596 130199: contig of 6604 bp in length

130200 130299: gap of 100 bp

130300 132629: contig of 2230 bp in length

132630 132728: gap of 100 bp

132730 135477: contig of 2748 bp in length

135478 135577: gap of 100 bp

135578 139010: contig of 3433 bp in length

139011 139110: gap of 100 bp

139111 141568: contig of 2458 bp in length

141569 141668: gap of 100 bp in length

142232 142131: contig of 463 bp in length

142132 142231: gap of 100 bp

142232 144866: contig of 2635 bp in length

144867 144966: gap of 100 bp

144967 148414: contig of 3448 bp in length

148415 148514: gap of 100 bp

148515 150504: contig of 1990 bp in length

150505 150604: gap of 100 bp

150605 151939: contig of 1355 bp in length

151960 152059: gap of 100 bp

152060 153455: contig of 1396 bp in length

153456 153555: gap of 100 bp

153556 155251: contig of 1666 bp in length

155252 155351: gap of 100 bp

155352 155869: contig of 1518 bp in length.

Location/Qualifiers

1. 156869

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="11"

/map="11q"

/clone="RP11-122H4"

1. 16262

/note="assembly-fragment"

16363. 32533

/note="assembly-fragment"

32634. 47033

```
misc_feature /note="assembly_fragment"
47134..60266
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
60367..70550
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
70651..83136
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
83237..90994
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
91095..99344
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
99445..106022
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
106123..112604
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
112705..118274
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
118375..123495
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
123596..130199
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
130300..132629
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
132730..135477
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
135578..139010
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
139111..141568
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
141669..142131
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
142232..144866
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
144967..148419
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
148515..150504
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
150605..151959
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
152060..153455
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
153556..155251
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
155352..156869
/note="assembly_fragment"
BASE COUNT 42936 a 35209 c 34104 g 42199 t 2421 others
ORIGIN
a)gment_scores:
Quality: 1202.00 Length: 1059
Ratio: 3.588 Gaps: 12
Percent Similarity: 31.634 Percent Identity: 31.539
Alignment block:
US-09-653-961-2 x AP002888
Align seg 1/1 to: AP002888 from: 1 to: 156869
19 ProArvYal.....AlacIYalProGlycIuaIacIugInProAlaPr 33
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
145244 CCGCGCTCCCTCCCGAGGTGTGCGGAGAGCTGAGCAGCCTCGCC 145293
33 OGIuIeUaIglUaIglUaIglYserThrAlaIeUeUlyscYsGIYL 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
145294 TGAGCTGTGAGGTGAGAGTGGAGCAGCAGCCCTTGTGAAGTGGGCC 145343
50 eUSeGIseGIInglYAsnIeUSeH1sValAsPTr..PhaSeT.... 64
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
145344 TCTCCAGTCCCAAGCAAGCTCAGCAGTGTGACTGTTCGTGAGT 145393
64 ..... 64
```

```
145394 GCTGGGCTCTTCAGAGAGAGCCCGGAGGAGAGACTCTGCCCTCTTG 145443
65 .....ValH1sLY 67
145444 TAGCCAGAGACCTTCTCTCTCTGTGTCTCCCTCATTAAGTCCACA 145493
67 sGIuYsArGThrIleUlePhaArGValArGInglYInglYInseRg 84
145494 GGAGAAAGCGAGCTCATCTTCCGTGTGGCCAGAGGCCAGAGCCAGACG 145543
84 IuPProGlyGIuYrGIuGIuArIleUSeIeUInAsPArGIYAlaPhr 100
145544 AACCTGGAGTACAGAGCGGCTCAGCCTCCAGACAGAGAGGCTCTCT 145593
101 IeUaIleUThrGIuValThrProGInAsPGIuArGIlePhaIeUcYsGI 117
145594 CTGGCTGTACTAGTCAAGTCAACCCCAAGACAGAGCTATCTTGTGCA 145643
117 nGIYsArGProArGSeRGIuYrArGIleGIuIeUaRGIYr. 133
145644 GGNCAAGCGCCCTCGTCCAGAGTACCGATCCAGTCCCGCTCA 145693
133 ..... 133
145694 GTGGTGACCTGTGTGTGACAGAGGTAGAGGTGTGAGAGTGG 145743
134 .....IYsAlAProGIuGIuPr 139
145744 CCAGATGTCCCTCATTTCTATTCTGTGAGAGCT..CGGAGAGAGC 145792
139 oAnIleGIuYsArProlaUeGIYIleProValAsnSeRyGIuPProG 156
145793 AAACATCCAGGTCAACCCCTGGGATCCCTGTGAACATAGAGAGCTG 145842
156 IuGIu..... 157
145843 AGGAGGTAGATGTGTGAGAGTGGCTGGGCACTGTGTCCAGCCGCT 145892
157 ..... 157
145893 GAGATGGGCTTGAAAGGGCTCGAGAGAGGAGAGACTCACCCGTGTC 145942
158 .....ValAlaThrCYsValGIYArGAsnGIYrProIleProGInV 172
145943 TCTTTAGTGTGCTACCTGTGTAGAGAGAGAGGATACCCATTTCTCA 145992
172 aIleTrPThrYsAsnGIYArProlaUeGIuGIuYsAsnArG... 187
145993 TCATCTGTATCAAGAAATGGCCGGCTCTGAAGAGAGAGAAAGCTGAG 146042
187 ..... 187
146043 TAGCCCTTCTCCAAAAGCCACCCCTGGGCCAGAGTGTGTAGCTCAAC 146092
187 ..... 187
146093 CTGTAAATCTGGTGTCTTGGAGGCCAAGCAGAGAGATCTGTGAGGTG 146142
187 ..... 187
146143 AGGAGTTCAGACCACTTGGACACAATAGAGTACCCTGTCTACAGA 146192
187 ..... 187
146193 AATATAAAAAATTAGCAGAGCATGTGTGTGTGTGCTGTGTCCAGATA 146242
187 ..... 187
146243 CTGGAGAGGTGAGGAGAGAGATCATATGAGCCAGAAATGAGAGCTG 146292
187 ..... 187
146293 CAGTGAAGTGAAGATTGCACACACACTCCGGGCTGTGGTGAAGAGCAG 146342
```

187 187
146343 ACCCATCTCTTAAAAAGAGCTGCTATGCTACTACACAGAACTGCC 146392
187 187
146393 TTATATTTTCATTAGCTTTTAACTTAAAAAATTATACAAATGCTACTT 146442
187 187
146443 GAAATATTTTCATGAGAAATTAAACGACAAAAAATCTAAGATTCTT 146492
187 187
146493 CTTCCTGAGACCCCTCCCGAATAATCTCTGCTATCAGGGTGTGT 146542
187 187
146543 TCTTTCAGCCCTATTCTATGTATTTGCTCATATATAGAAATATTCTAG 146592
187 187
146593 AATGATATAGGCTTCTGTGTATTATCTAAATCAGTCATTCTTAACA 146642
187 187
146643 GGGGTGATTTTGTACCCCTCTCTCTAGAGATCTGGCAATGTGTGA 146692
187 187
146693 GATATTTTGGTGTCAACATAGAGGGGGTCTACTCCATCTTAAGT 146742
187 187
146743 TAGAGAGCAAGATGTGCTAATCTTATAGGACAGACAGGCC 146792
187 187
146793 CACATATTAAGATTCAGCTGTGCTTAAACATCAGTAGTGTGCTGGC 146842
187 187
146843 TCACGCTGTATATCCACACTTTTGGAGCCAAAGTGGGGATCACCT 146892
187 187
146893 GAGTGGGAGTACAAAGACAGCTGTCCACAGAGAAACCCATCTC 146942
187 187
146943 TACTGAAATACAAAAGTAGCCGGCGGTGGCCATGTCTGTATGCC 146992
187 187
146993 AGCTACTGAGAGCTGAGAGAGAAATCATTGAAACCCGGAGGAGG 147042
187 187
147043 TGGAGTTTCGGTGAAGCCAGATTTACCCCTGCACCTCAGCTGGGCCA 147092
187 187
147093 CAAGAGTGAACCTTTTAAAAAATAAATAATATCAGTAGTG 147142
187 187
147143 CTGAGAAACCTGTCTAAGTGTGTATGATACATTTGTTAGACAT 147192
187 187
147193 TTCTTTATACAATGTTCTGGTCACTGATTTAGATCAACTGATGTT 147242

187 187
147243 TTGCTTACTGCCAAGTTTCCATCTACTACGATAGCAGTAGTCGAGTTCA 147292
187 187
147293 CCATCCCATTTAGTGCATCTAGACGGCTGCTGTTTATCATTTGC 147342
187 187
147343 AGCATCTTTGCACACATCTTGATATATGACGACATGAAATGTTTTT 147392
187 187
147393 CTAGGTTGACACTGACGATAAAGTGTGGTTGTAAGGTTTCCAGCT 147442
187 187
147443 TGCATTTGTAAGCTGGCTTTCACAGGGGACAGGGGCTATTATAGTGC 147492
187 187
147493 CCTGCCAACCCCACTGGACAAACCTTAGGGGGCTGGAGTGGGCTG 147542
188 ValHisIleGlnse 192
147543 AGGTGAGTCTTCTCCCTTCTCCCTGCCAGGGGCTCACATTCAGTC 147592
192 rSerGlnThrValIleGlnSerGlyLeuTyrThrLeuGlnSerIleLeu 209
147593 GTCCAGACTGTGGAGTGCAGTGGTTTGACACTTGCAGAGTATCTGA 147642
209 ysaIaGlnLeuValIleGlnPlyAspAlaIlePheTyrCysGlnLeu 225
147643 AGCAGACGCTGGTTAAAGAACAAATATCCAGTTTACTGTCGACTC 147692
226 AsnTyrArgLeuProSerGlyAsnHisMetIleGlnSerArgGlnVal 242
147693 AACTACGGGTGCCAGTGGAGACCATATGAAAGAGTCCAGGAGTGC 147742
242 rValProValPheTyr..... 247
147743 CGTCCGTGTTTCTGTGAGTACTAACCTGACTCTGCTACTGGGCTG 147792
247 247
147793 GGGCTGTGGAACCTCTGTGCTCCCTTATCCAGCCAGGGTTGAC 147842
248 ProThrGlnIleVal 252
147843 TCATGGGCCAAGAGTACCCCTTGTCCCTATGCAGACCCGAAAAAGT 147892
252 LTrPLeuGlnValIleGlnProValIleGlnLeuIleGlnAspArgVal 269
147893 GTGGCTGGAAGTGGAGCCCGTGGGAATCTGAAGAGAGGAGCCCGTGG 147942
269 LuIleArgCysIleuAlaAspIleAspProPheProHisPheSerIleSer 285
147943 AATCAGGTGTTTGGCTATGSCAACCTCCACCACTTCAGCATCGC 147992
286 Lys..... 286
147993 AAGCAGGTGGGGGCTTGTGCTGGGTGCAGGTGGGATGCACAGAG 148042
287 G 287
148043 ACCCAGGTGAGTGGCGGAGGTGCACCTGTCACCTCCCAATCTCC 148092
287 In..AsnProSerThrArgGlnAlaIleGlnIleGlnIleThrThrAsnAspArg 303
148093 CTAGAACCCCGACGACGAGGAGGAGGAGAAAGAACCAACGACGAAAG 148142
303 IyalleuValleuGlnProAlaArgIleGlnHisSerGlyArgTyrGln 319

```

|||||
148143 GGCTCTGCTGCTGAGGCTGCCGGAAGAACACAGTGGCGCTATGAA 148192
320 CysGlnGlyLeuAspLeuAspThrMetLysSerLeuSerGluProG1 336
|||||
148193 TGTCAAGGCGCTGACTTGACACCATGATATCGCTGAGTGAACACA 148242
336 nGluLeuLeuValAsn..... 341
|||||
148243 GGAAGTACTGTTGACACTGTGAGGGGCTGGGAGCCAGGACAGGGGACCA 148292
341 ..... 341
148293 GGCTGGGCAATAGGGAAGTGGCCACCTGCGCTGGGTGACATGATG 148342
342 .....Tyr..ValSerAspValArgValSerProLala1a.ProG1 354
|||||
148343 CTGTACCCCCAGATGTGTCTGACGTCCGAGTGAAGTCCCGCAGCAGACTGA 148392
354 uArgGlnGluGlySerSerLeu 361
|||||
148393 GAGACAGGAAGGACAGCGCTC 148414

```


alignment block:
US-09-653-961-2 x AA220930 ..

Align seg 1/1 to: AA220930 from: 1 to: 1950

```
1 MetGlyLeuProAlaGlyLeuValCysAlaPheLeuLeuAlaAlaCysCysCys 17
1 ATGGGGCTTCCAGCTGGTCTGCGCTTTCCTTGGCTCCGCGCTCGCTG 50
17 scyProAlaGlyAlaGlyValProGlyGluAlaGluGlnProAlaProG 34
51 CTGTCTCGCGCTCGGGGTGTGCCGGAGAGCTGAGCAGCTCGCGCTG 100
34 LuLeuValGluValGluValGlySerThrAlaLeuLeuLysCysGlyLeu 50
101 AGCTGGTGGAGTGGAGAGTGGGACAGCACAGCCCTTGTGAAGTGGCGCTC 150
51 SerGlnSerGlnGlyAsnLeuSerHisValAspTrpPheSerValHisLys 67
151 TCCAGTCCCAAGGCAACCTCAGCATCTCCAGCTGGTTTCTGTCCACAA 200
67 sGluLysArgThrLeuLeuIlePheArgValArgGlnGlyGlnGlnSerG 84
201 GGAGAAAGCGGAGCGCTCATCTTCCGTGCGCCAGGCGCAGGGCCAGAGCG 250
84 LuProGlyGluTyrGluGlnArgLeuSerLeuGlnAspArgGlyValAlaThr 100
251 AACCTGGGAGTACGACGAGCGGCTCAGCCTTCACGACAGACAGAGGGCTACT 300
101 LeuAlaLeuThrGlnValThrProGlnAspGluArgIlePheLeuCysG 117
301 CTGGGCCCTGACATCAAGTCAACCCCAAGACGAGCGCATCTTCTGTGCCA 350
117 nGlyLysArgProArgSerGlnGluTyrArgIleGlnLeuArgValTyrL 134
351 GGGCAAGCGCCCTCGTCCAGAGTACCGCATCCAGCTCGCGCTGTACA 400
134 ValAlaProGluGluProAsnIleGlnValAsnProLeuGlyIleProVal 150
401 AAGCTCCGGAGAGGCCAAACATCCAGGTCAACCCCTGGCGCATCCCTGTG 450
151 AsnSerLysGluProGluGluValAlaThrCysValGlyArgAsnGlyTyr 167
451 AACAGTAAGAGAGCTGAGAGAGTCCGTACTGTGTAGGAGGAAACGGGTA 500
167 ProIleProGlnValIleTyrTyrLysAsnGlyArgProLeuLysGluG 184
501 CCCCATTTCTCAAGTCATCTGGTACAAAGATGGCGCGCTCTGAAGAGAGG 550
184 LuLysAsnArgValHisIleGlnSerSerGlnThrValGluSerSerGly 200
551 AGAAGAAGCGGGTCCCATTCAGTGTGCTCCACACTGTGAGTGCAGTGTG 600
201 LeuTyrThrLeuGlnSerIleLeuLysAlaGlnLeuValLysGluAspLys 217
601 TTGTACACCTTGCAGATATCTGAAAGGACAGCTGGTTAAAGAAAGACAA 650
217 sAspAlaGlnPheTyrCysGluLeuAsnTyrArgLeuProSerGlyAsnH 234
651 AGATGGCCCAAGTTTACTGTGAGCTCAACTACCGGCTGCCAGTGGGAACC 700
234 IsMetLysGluSerArgGluValThrValProValPheTyrProThrGlu 250
701 ACATGAAGGAGTCCAGGAAAGTACACGCTCCCTGTTTCTACCCGACAGAA 750
251 LysValTyrPleuGluValGluProValIleGlyMetLeuLysGluGlyAsp 267
751 AAAGTGTGGCTGGAAAGTGGAGCCGTGGGAATGCTGAAGGAAGGGGACCG 800
267 gValGluIleArgCysLeuAlaAspGlyAsnProProHisPheSerI 284
801 CGTGGAAATCAAGGTGTTGGCTGATGGCAACCTCCACCACTTACAGA 850
```

```
284 LeSerLysGlnAsnProSerThrArgGluAlaGluGluGlnThrThrAsn 300
851 TCAGACAGCAGAACCCACAGCACAGGAGGACAGAGAACAGAACACAC 900
301 AspAsnGlyValLeuValLeuGluProAlaArgLysGluHisSerGlyAr 317
901 GACACAGGGGTCTGTGTGTGGAGCTGCCCGGAAGAACACAGTGGCGG 950
317 gTyrGluCysGlnGlyLeuAspLeuAspThrMetIleSerLeuLeuSerG 334
951 CTATTGAATGTCAAGGCGCTGAGCTTGGACACCATGATATGCTGTGAGTG 1000
334 LuProGlnGluLeuLeuValAsnTyrValSerAspValArgValSerPro 350
1001 AACACAGAGAACTACTGTGTAATACTATGTGTGAGCTCCGAGTGAATCC 1050
351 AlaAlaProGluArgGlnGlySerSerLeuThrLeuThrCysGluAla 367
1051 GCAGCCCTGAGAGACAGGAAGGACAGACAGCTCACCCCTGACGTGAGGC 1100
367 ArgLysSerGlnAspLeuGluProPheGlnTrpLeuArgGluThrGlyG 384
1101 AGAGAGTACGACAGACCTCGAGTTCCAGTGGCTGACAGAAAGACAGGCC 1150
384 LuValLeuGluArgGlyProValLeuGlnLeuHisAspLeuLysArgGlu 400
1151 AGGTGCTGAAAGGGGCGCTGTGCTTCAGTTGCATGACCTGAACCGGAG 1200
401 AlaGlyLysGlyTyrArgCysValAlaSerValProSerIleProGlyLe 417
1201 GCAGAGCGCGGCTATCGCTCGTGGGTGTGTGCCAGCATACCCGAGACT 1250
417 uAsnArgThrGlnLeuValAsnValAlaIlePheGlyProProThrPheMe 434
1251 GAACCGACACACAGTGTGTCAACGTGGCCATTTTGGCCCCCTTGGATGG 1300
434 IapheLysGluArgLysValTyrValLysGluAsnMetValLeuAsnLeu 450
1301 CATTCAAAGAGAGAGAGAGTGTGGTGAAGAAATATGATGTTGAATCTG 1350
451 SerCysGluLysSerGlyHisProArgProThrIleSerTrpAsnValAs 467
1351 TCTTGTGAAGCTTAGGGGACACCCCGGCCACCATCTCTCGGAACGTCAA 1400
467 nGlyThrAlaSerGluGlnAspGlnAspProGlnArgValLeuSerThrL 484
1401 CGGACAGGCAAGTGAACACAGCAAGATCCACAGCGAGTCTGAGCAGACC 1450
484 euAsnValLeuValThrProGluLeuGlnGluThrGlyValGluCysThr 500
1451 TGAATGTCTCTGTGACCCCGGAGCTGTTGAGACAGGTGTGAATGTGACG 1500
501 AlaSerAsnAspLeuGlyLysAsnThrSerIleLeuPheLeuGluLeuVa 517
1501 GCCTCCAAACGACCTGGGGCAAAAACACACAGCATCTCTTCTGGAGGTGT 1550
517 LasnLeuThrThrLeuThrProAspSerAsnThrThrThrGlyLeuSerT 534
1551 CAATTTAAACACCTCTCACACAGACTCCAAACACACAGCTGGCTTCAGCA 1600
534 hrSerThrAlaSerProHisThrArgAlaAsnSerThrSerThrGluArg 550
1601 CTTCACAGTCCAGTCTCTATACAGAGCCAAACAGACCTCCACAGAGAGA 1650
551 LysLeuProGluProGluSerArgGlyValValIleValAlaValIleVa 567
1651 AAGCTGCCGGAGCCGGAGAGCCGGGCGGTGTATCTGTGGCTGTGATGTG 1700
567 lCysIleLeuValLeuAlaValLeuGlyAlaValLeuTyrPheLeuTyrL 584
1701 GTGCATCTCTGTGCTGGGCTGTGGGCTGTGCTCTATTTCTCTGTATA 1750
```



```
584 yslYsgLYLysLeuProCysArGArGserGlyLysGlnGluIleThrLeu 600
1751 AGAAGGGCAAGCTGCGCTGACAGCGCTCAGGGAAGCAGAGATCACGCTG 1800
601 ProProSerArgLYSerSerGluLeuValValGluVallySerSerAspLYSLe 617
1801 CCCCCGTCGTCAAGAGCGAACTGTGTAGTGAAGTCAATTAACCT 1850
617 uProGluGluMetGlyLeuLeuGlnGlySerSerGlyAspLYSArgAlaP 634
1851 CCCAAGAGAGATGGGCTCTCTGACAGGACGACGCGGTACACAGAGGCGCTC 1900
634 roGlyAspGlnGlyGluLYSTyriLeaAspLeuArgHis 646
1901 CGGAGACACGAGGAGAGAAATATCATCATCTGAGGCAAT 1938
seq_name: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAZ20931
seq_documentation_block:
ID_AAZ20931 standard; cDNA; 1960 BP.
XX
AC_AAZ20931;
XX
DT 01-DEC-1999 (first entry)
XX
DE Human MUC18 cDNA sequence, previously published (Johnson 1989).
XX
KM prostate cancer; melanoma; cell adhesion; glycoprotein;
XX metastasis; treatment; detection; diagnostic test; ds.
XX Homo sapiens.
OS
FH Key Location/Qualifiers
FT CDS 8..1948
FT /tag= a
FT /product= "Human MUC18"
FT /note= "human MUC18"
XX
PN MO9945392-A1.
XX
PD 10-SEP-1999.
XX
PF 02-MAR-1999; 99MO-US04850.
XX
PR 03-MAR-1998; 98US-0076664.
XX
PA (UYEM-) UNIV EMORY.
XX
Wu G;
XX
PI WPI; 1999-540899/45.
XX
DR P-PSDB; AAY42405.
XX
PT Detection of metastatic prostate cancer, by detection of MUC18
XX expression in prostate cancer cells.
XX
PS Disclosure; Page 8; 80pp; English.
XX
XX This is the previously published nucleotide sequence of the Human MUC18
XX cell adhesion glycoprotein, which is expressed on the surface of
XX melanoma cells, and can be used as a marker for prostate cancer.
XX This nucleotide sequence is virtually identical to the huMUC18 sequence
XX set forth in the invention, except for seven amino acid residues (which
XX may be due to allelic differences). However the amino acid sequence of
XX Johnson's sequence was 43 residues shorter than the sequence proposed
XX by the inventors.
XX The presence of this glycoprotein has been correlated with the ability of
XX melanomas to metastasize. MUC18 is also associated with normal vascular
XX tissue, and on the smooth muscle of venules, and it expresses
XX sporadically on capillary epithelium.
XX The method can be used as a diagnostic test for prostate cancer which
XX has a relatively high potential for metastasis or which has metastasized.
XX The physician can then choose the appropriate surgical, chemotherapeutic
```

```
CC or radiation treatment regimens. In addition, antibody specific to MUC18
CC can be used to prevent metastasis of Prostate Cancer Cells.
XX
SQ Sequence 1960 BP; 457 A; 557 C; 597 G; 349 T; 0 other;

alignment_scores:
Quality: 3327.00 Length: 646
Ratio: 5.166 Gaps: 0
Percent Similarity: 99.690 Percent Identity: 98.916

alignment_block:
US-09-653-961-2 x AAZ20931 ..

Align seg 1/1 to: AAZ20931 from: 1 to: 1960

1 MetGlyLeuProArGLeuValCysAlaPheLeuLeuAlaAlaCysCysCy 17
|||||
8 ATGGGGCTTCCAGGCTGGTCTGCGCTTCTTGCTGCGCGCTGCTG 57
|||||
17 SCysProArGValAlaGlyAlaProGlyGluAlaGluGlnProAlaProG 34
|||||
58 CTGTCTCTCGCGTGGCGGGTGTGCCGAGAGGCTGAGAGCCTGGCGCTG 107
|||||
34 LuLeuValGluValGluValGlySerThrAlaLeuLeuLYSCysGlyLeu 50
|||||
108 AGCTGTGGAGGTGGAAAGTGGGACACAGCCCTTTGTAAGTGGGCTTC 157
|||||
51 SerGlnSerGlnGlyAsnLeuSerHisValAspTyrPheSerValHisLy 67
|||||
158 TCCACAGTCCCAAGCAACCTCAGCCATGTGCACTGTGTTCTTGTCCACA 207
|||||
67 sGlyLYSArgThrLeuIlePheArgValArgGlnGlnGlnSerG 84
|||||
208 GGAGAAAGCGGACGCTCATCTTCGTTGTGGCCAGGCGCCAGGCGAAGG 257
|||||
84 LuProGlyGluTyrGlnGlnArgLeuSerLeuGlnAspArgGlyAlaThr 100
|||||
258 AACCTGGGAGGTACGACAGCGGCTCAGCCTCCAGACAGAGGGCTACT 307
|||||
101 LeuAlaLeuThrGlnValThrProGlnAspGluArgIlePheLeuCysG 117
|||||
308 CTGGCCCTGACTCAAGTCAACCCCCAAGACGAGCGCATCTTCTTGAGCA 357
|||||
117 nGlyLYSArgProArGSerGlnGluTyrArgIleGlnLeuAlaGlyTyrL 134
|||||
358 GGGCAAGCGCGCTCGGTCCAGAGATCCGATCCAGCTCCGCTCTACA 407
|||||
134 YsaAlaProGluGluProAsnIleGlnValAsnProLeuGlyIleProVal 150
|||||
408 AACCTCGGAGAGGCCAAACATCCAGGTCAACCCCTGGGCAATCCCTGTG 457
|||||
151 AsnSerLYSGluProGluGluValAlaThrCysValGlyArgAsnGlyTy 167
|||||
458 AACAGTAAGAGCGCTGAGAGGTCGCTACCTGTGTGAGGAGAACGGGTA 507
|||||
167 rProIleProGlnValIleTyrTyrLYSAsnGlyArgProLeuLYSGluG 184
|||||
508 CCCCATTCCTCAAGTATCTGTGTACAAAGATGGCGGCTGTGAAGGAGG 557
|||||
184 LuLYSAsnArgValHisIleGlnSerSerGlnThrValGluSerSerGly 200
|||||
558 AGAAGAAACCGGGTCCACATTCAGTCTCTCCACAGCTGTGAGTCAAGTGT 607
|||||
201 LeuTyrThrLeuGlnSerIleLeuLYSAlaGlnLeuValLYSGluAspLy 217
|||||
608 TTGTACACCTTGCAGAGTATCTGAAGGCACAGCTGTGTAAGAAGACAA 657
|||||
217 sAspAlaGlnPheTyrCysGluLeuAsnTyrArgLeuProSerGlyAsnH 234
|||||
658 AGATGGCCAGTCTTACTGTGAGCTCAACTACCGGCTGCCCAATGGGAACC 707
|||||
234 IsMetLYSGluSerArgGluValThrValProValPheTyrProThrGlu 250
```

```
|||||
708 ACATGAAGAGTCACAGGAAGTCACCGTCCCTGTTTCTACCCGACAGAA 757
251 LysValThrPLeuGluValGluProValGlyMetLeuLysGluGlyAsp 267
758 AAAGTGTGCTGGAAAGTGGAGCCCGGGAAATGCTCAAGGAAGGGGACCG 807
267 GValGluIleArgCysLeuAlaAspGlyAsnProProOHISpHeserI 284
808 CGGGAAATCAGGTGTTTGGCTGATGGCAACCTCCACACACTTCAGCA 857
284 LeSerLysGlnAsnProSerThrArgGluAlaGluGluThrThrAsn 300
858 TCAGCAAGCAGAACCCACAGCAGGAGGAGAGAGAGAACAAACAAAC 907
301 AspAsnGlyValLeuValLeuGluProAlaArgLysGluHisSerGly 317
908 GACAAAGGGGTCTGCTGGAGCTGCGGAAAGAACACAGAGTGGCG 957
317 gTyrgLysGlnGlyLeuAspLeuAspThrMetIleSerLeuLeuSerG 334
958 CTATGATGTGCAAGCCTGGAGACTGGACACATGATGCTGCTGAGTG 1007
334 LueProGlnGluLeuValAsnTyrValSerAspValArgValSerPro 350
1008 AACCAAGAGAACTAGTGAATGATGTGTGACGTCCGAGTGAAGTCCC 1057
351 AlaAlaProGlnArgGlnGluGlySerSerLeuThrLeuThrCysGln 367
1058 GCAGCCCTGAGAGACAGAGAGGAGCAGCAGCTCACCTGAGTGAAGC 1107
367 agLueSerSerGlnAspLeuGluPheGlnTrpLeuArgGluGluThrGly 384
1108 AGAGAGTAGCCAGACCTGAGTTCAGTGGCTGAGAGAAAGAGACGACC 1157
384 LueValLeuGluArgGlyProValLeuGlnLeuHisAspLeuLysArgGlu 400
1158 AGGTGCTGGAAAGGGGCGCTGTGCTTCAGTGCATGACCTGAAGCGGAG 1207
401 AlaGlyGlyGlyTyrArgCysValAlaSerValProSerIleProGly 417
1208 GCAGGAGGGGCTATGCTGCTGCGTGGCTGTGCCACGATACCGGCT 1257
417 uAsnArgThrGlnLeuValAsnValAlaIlePheGlyProProTrpMet 434
1258 GAACCGGACACAGCTGTGACGTCGCAATTTTGGCCCCCTTGGATGG 1307
434 LabeledLysGluArgLysValThrValLysGluAsnMetValLeuAsnLeu 450
1308 CATTCAAGGAGAGAGAGTGTGGTGAAGAGAAATGTGTGAATCTG 1357
451 SerCysGluAlaSerGlyHisProArgProThrIleSerTrpAsnVal 467
1358 TCTTTGAAAGCTGACGGACCCCGCCGACCATCTCTCGAGACTCA 1407
467 ngLysThrAlaSerGluGlnAspGlnAspProGlnArgValLeuSerThr 484
1408 CGGACGCGCAATGAAACAAACCAAGATCCACAGAGTCTGACACCC 1457
484 euAsnValLeuValThrProGluLeuLeuGluThrGlyValGlyCysThr 500
1458 TGAATGCTCTGAGACCCCGGAGCTGTGAGACAGGTGTGAATCTCAG 1507
501 AlSerAsnAspLeuGlyLysAsnThrSerIleLeuPheLeuGluLeuVal 517
1508 GCCTCCACACACCTGGGCAAAAACACAGATCTCTTCTCGAGCTGGT 1557
517 LAsnLeuThrThrLeuThrProAspSerAsnThrThrGlyLeuSerT 534
1558 CAATTTAAACACCTCCACACACAGACTCCAAACACACACAGCTGCTCAGCA 1607
534 hSerThrAlaSerProHisThrArgAlaAsnSerThrSerThrGluArg 550
|||||
```

```
1608 CTTCACATGCCATGCTCATACAGAGCCAAACAGACCTCCACAGAGAGA 1657
551 LysLeuProGluProGluSerArgGlyValValIleValAlaValIleVal 567
1658 AAGCTGCCGAGACCCGAGAGCCGGGGCTGGTCATCTGCTGATGATGT 1707
567 LysIleLeuValLeuAlaValLeuGlyAlaValLeuTyrPheLeuTyrL 584
1708 GTGCATCTGTGCTGCTGGGGGTGTGGGGCTGTCTCTATTTCTCTATA 1757
584 yLysGlyLysLeuProCysArgArgSerGlyLysGlnGluIleThrLeu 600
1758 AGAAGGGCAAGCTCCGCTGACGGCTCAGGGAAGCAGAGATCACGCTG 1807
601 ProProSerArgLysSerGluLeuValValGluValLysSerAspLysLe 617
1808 CCCCCGTCTGTAGACCGAACTGTGATTAAGTCAAGTAAAGCT 1857
617 uProGluGluMetGlyLeuLeuGlnGlySerSerGlyAspLysArgAlaP 634
1858 CCCAGAAAGATGGGCTCTCTGACAGGCGCAGCAGCGGTGACAAAGGGCTC 1907
634 roGlyAspGlnGlyGluLysTyrIleAspLeuArgHis 646
1908 CGGGAGACGAGGAGAGAAATATCATCATGAGGCAAT 1945
seq_name: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH02921
seq_documentation_block:
ID AAH02921 standard; DNA: 3293 BP.
XX
AC AAH02921:
XX
DE 15-JUN-2001 (first entry)
XX
DE Human shear stress-response coding sequence SEQ ID NO: 95.
XX
KW Human; shear stress-response protein; vascular disease;
XX arteriosclerosis; ds.
XX
OS Homo sapiens.
XX
PN W0200125427-A1.
XX
PD 12-APR-2001.
XX
PE 02-OCT-2000; 2000MO-JP06840.
XX
PR 01-OCT-1999; 99JP-0280976.
XX
PA (RYOW ) KYOMA HAKKO KOGYO KK.
XX
PA (NOJI/) NOJIMA H.
XX
PI Nojima H, Yoshisue H, Obayashi M, Ota T, Kawabata A, Sakurada K,
PI Kuga T, Sekine S, Nakamura Y, Sugano S,
XX
DR WPI: 2001-266308/27.
XX
P-PSDB: AAB90798.
XX
PT DNA sequences, proteins encoded by them and antibodies against them
PT useful in diagnosis and treatment of vascular disease caused by
PT arteriosclerosis -
XX
PS Claim 20; Page 491-496; 678bp; Japanese.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human shear stress response proteins. These are useful in the
CC diagnosis, treatment and screening of vascular diseases caused by
CC arteriosclerosis, including heart failure, post-PTCA restenosis and
CC hypertension.
XX
SQ Sequence 3293 BP; 786 A; 916 C; 924 G; 667 T; 0 other;
```

Mon Jul 1 11:07:41 2002

us-09-653-961-2.ring

Page 5

alignment_scores:

Quality: 3327.00 Length: 646
Ratio: 5.166 Gaps: 0
Percent Similarity: 99.690 Percent Identity: 98.916

alignment_block:
us-09-653-961-2 x AAH02921

Align seg 1/1 to: AAH02921 from: 1 to: 3293

```
1 MetGlyLeuProAlaGlyLeuValCysAlaPheLeuAlaAlaCysCysCy 17
   |||||
8 ATGGGGCTTCCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 57
   |||||
17 sCysProAlaGlyAlaAlaGlyValProGlyGluAlaGluInProAlaPro 34
   |||||
58 CTGTCTCCGCGCGGGGGTGTGCCGGAGAGCTGAGCAGCGCTGCGGCTG 107
   |||||
34 LuleuValGluValGluValGlySerThrAlaLeuLeuLysCysGlyLeu 50
   |||||
108 AGCTGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAG 157
   |||||
51 SerGlnSerGlnGlyAsnLeuSerHisValAspTrpPheSerValHisLy 67
   |||||
158 TCCTCCAGTCCCAAGGCAACCTCAGCCATGTCGACTGCTTTCTGTCACAA 207
   |||||
67 sGluysArGThrLeuLlePheArGValArGInGlyInGlyInSerG 84
   |||||
208 GGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 257
   |||||
84 LurProGlyGlyTrpGlnArgLeuSerLeuGlnAspArgGlyValAlaThr 100
   |||||
258 AACCTGGGAGGTACGACAGCGGCTCAGCTTCAGSACAGAGGGGCTACT 307
   |||||
101 LeuAlaLeuThrGlnValThrProGlnAspGluArgGlyLeuLeuCysG 117
   |||||
308 CTGGCCCTGACTCAAGTACACCCCAAGAGAGAGCGATCTCTGTGTGCA 357
   |||||
117 nGlyysArGProAlaGserGlnGlyTrpArgGlyLeuValArgValTyr 134
   |||||
358 GGGCAAGGCGCCCTGCTGCCAGAGATACGCAATCCAGCTCGCGCTACAA 407
   |||||
134 ysaAlaProGlyGluProAsnLleGlnValAsnProLeuGlyLleProVal 150
   |||||
408 AAGCTCCGAGGAGGAGCAACATCAGGTCAACCCCTGGGATCCCTGTG 457
   |||||
151 AsnSerLysGluProGlnGluValAlaThrCysValGlyArgAsnGlyTy 167
   |||||
458 AACAGTAAGAGACCTGAGAGAGTGGCTACCTGTGTAGGAGAGACGGGTA 507
   |||||
167 rProLleProGlnValLleTrpTrpLysAsnGlyArgProLeuLysGlu 184
   |||||
508 CCCCAATCTCAAGTCACTGTGTACAAAGATGGCGGCTCTGAGAGAGG 557
   |||||
184 LuleysAsnArgValHisLleGlnSerSerGlnThrValGlyUserSerGly 200
   |||||
558 AGAAGAACCGGGTCCACATTCAGTCCGAGGCTGTGAGTGCAGATGTGT 607
   |||||
201 LeuTrpThrLeuGlnSerLleLeuLysAlaGlnLeuValLysGluAspLy 217
   |||||
608 TTCTACACCTTCCAGAGTATTCGTAAAGGACAGCTGTGTTAAAGAACAA 657
   |||||
217 sAspAlaGlnPheTrpCysGluLeuAsnTrpArgLeuProSerGlyAsn 234
   |||||
658 AGATGGCCAGTGTTCGTGAGCTCAACATACGGGCTGCCAGTGGAGAAC 707
   |||||
234 IsMetLysGlnSerArgGluValThrValProValPheTrpProThrGlu 250
   |||||
708 ACATGAGAGAGTCCAGGAGATCACGCTCCTGTTTCTACCCGACAGAA 757
   |||||
251 LysValTrpLeuGluValGluProValGlyMetLeuLysGluGlyAspAr 267
   |||||
```

```
758 AAAGTGTGGCTGSAAGTGGAGCCCGTGGGAAATGCTGAAGAGAGGAGCAG 807
267 gValGluLleArgCysLeuAlaAspLysAsnProProHisPheSer 284
   |||||
808 CGTGAATATCAGTGTGTGCTGATGGACACCTCCACACACTCAGCA 857
   |||||
284 LLeSerLysGlnAsnProSerThrArgGluAlaGluGluGluThrThrAsn 300
   |||||
858 TCAGCAAGCAAGAACCCCGACCGAGGAGGAGGAGAGAGAGAGAGAGAGAG 907
   |||||
301 AspAsnGlyValLeuValLeuGluLurProAlaArgLysGluHisSerGlyAr 317
   |||||
908 GACAAAGGGGTCCGTGGTGGAGCTGCCGGAAGAGAACAGTGGGGG 957
   |||||
317 gTyTrpLysGlnGlnLysAsnLeuAspThrPheLleSerLeuLeuSerG 334
   |||||
958 CTATGAATGTACGGCTGAGAACCTTGSACACATGATATCCGTCTGAGTG 1007
   |||||
334 LurProGlnGluLeuLeuValAsnTrpValSerAspValArgValSerPro 350
   |||||
1008 AACCAAGAACTACTGTGTAATGATGTGTGAGCTCCAGTGAATCC 1057
   |||||
351 AlaAlaProGlnLurArgGlnGlySerSerLeuThrLeuThrCysGluAl 367
   |||||
1058 GCAGCCCTGAGAGACAGAGAGGAGCAGACCTCAGCTGACCTGTGAGGC 1107
   |||||
367 sGluSerSerGlnAspLeuGluLurPheGlnTrpLeuArgGlyGluThrGly 384
   |||||
1108 AGAGAGTACGACGACCTCGATGTCCAGTGGGTGAGAGAGAGAGAGAGAC 1157
   |||||
384 LurValLeuGluLurArgGlyProValLeuGlnLeuHisAspLeuLysArgGlu 400
   |||||
1158 AGCTCTGTGAAGGGGCGCTGTCTTCACTTCATGATGATCTGAAGGGAG 1207
   |||||
401 AlaGlyGlyGlyTrpArgCysValAlaSerValProSerLleProGlyLe 417
   |||||
1208 GAGAGAGGCGGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1257
   |||||
417 uAsnArGThrGlnLeuValAsnValAlaLlePheGlyProProTrpMet 434
   |||||
1258 GACCGCACACACCTGCTCAAGCTGGCCATTTTGGCCCCCTGTGATGG 1307
   |||||
434 LArPheLysGluArgLysValTrpValLysGluAsnMetValLeuAsnLeu 450
   |||||
1308 CATTCAGAGAGAGAGTGTGGTGTAAAGAAATATGCTGTGATATC 1357
   |||||
451 SerCysGluAlaSerGlyHisProArgProThrLleSerTrpAsnValAs 467
   |||||
1358 TTTGTGAAAGGCTGAGGGCACCCCGGCCACCATCTCTGTGAACGTCAA 1407
   |||||
467 nGlyThrAlaSerGluGlnAspGlnAspProGlnArgValLeuSerThr 484
   |||||
1408 CGGCACGGCAAGTGAACAAACCAAGATCCACAGCGAGTCCAGACACC 1457
   |||||
484 euAsnValLeuValThrProGluLeuGluGluThrGlyValGlyCysThr 500
   |||||
1458 TGAATGCTCTGCTGACCCGAGACCTGTGGACAGAGGTGATGATGAGC 1507
   |||||
501 AlaSerAsnAspLeuGlyLysAsnThrSerLleLeuPheLeuGluLeu 517
   |||||
1508 GCTTCACAGACCTGGGCAAAACACACACATCTCTCGAGGGGGG 1557
   |||||
517 LAsnLeuThrThrLeuThrProAspSerAsnThrThrThrGlyLeuSer 534
   |||||
1558 CAATTTAACACCTCTCACACCACTCCAAACCAACACTGGCTGAGCA 1607
   |||||
534 hrSerThrAlaSerProHisThrArgAlaAsnSerThrSerThrGluArg 550
   |||||
1608 CTTCACATGCGCAGTCTCATACAGAGCCCAACAGACCTCCACAGAGAGA 1657
   |||||
551 LysLeuProGluLurProGlnSerArgGlyValAlaLleValAlaValIle 567
   |||||
1658 AAGCTGCCGAGGAGGAGAGCCGGGCGGTGTCATCGTGGCTGTGATGT 1707
   |||||
```

```
567 ICysIIeLeuValIleuAlaValIleuGlyAlaValIleuTyrrhLeuTyrl 584
|||||
1708 GTGCATCTGTCGTGCGGTGCGGTGTCCTCTATTCCTCTATTA 1757
|||||
584 yslYsgIyLysLeuProCysArGArGSerGlySgIngluIleThrLeu 600
|||||
1758 AGAAGGCAAGCTCCGCTGCGAGGCGCTCAGGAAAGCAAGATCAGCTG 1807
|||||
601 ProProSerArGlySerGlyLeuValIleuValIleuValIleuValIleu 617
|||||
1808 CCCCCTCTCTGTAAGACCAACTGTGTGAAGTTAAAGTCAGATAGCT 1857
|||||
617 uProGluGluMetGlyLeuLeuGlnGlySerSerGlyAspLysArGAlap 634
|||||
1858 CCCAGAAAGATGGGCTCTCTGACAGGCGACAGCGGTGACAAAGAGGCTC 1907
|||||
634 roGlyAspGlnGlyLysTyrrhLeuAspLeuArGHis 646
|||||
1908 CGGAGAGCCAGGAGAGAAATACATCGATCTGAGGCAAT 1945
|||||
seq_name: /STD55/gcgdata/geneseq/geneseqn-emb1/MH199.DAT:AAZ20932
seq_documentation_block:
ID AAZ20932 standard; cDNA; 1962 BP.
XX
XX AAZ20932;
AC
XX
XX 01-DEC-1999 (first entry)
XX
XX Human MUC18 cDNA sequence, modified to facilitate cloning.
XX
XX prostate cancer; melanoma; cell adhesion; glycoprotein; cloning;
KM metastasis; treatment; detection; diagnostic test; ds.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH 1..26
FT /*tag= a
FT /*note= "Primer Bf1"
FT primer_bind complement (1931..1955)
FT /*tag= b
FT /*note= "ER6A primer"
XX
XX MO9945392-A1.
XX
XX 10-SEP-1999.
XX
XX 02-MAR-1999; 99WO-US04850.
XX
XX 03-MAR-1998; 98US-0076664.
XX
XX (DYEM-) UNIV EMORY.
XX
XX Wu G;
XX
XX WPI: 1999-540899/45.
XX
XX Detection of metastatic prostate cancer, by detection of MUC18
XX expression in prostate cancer cells
XX
XX Disclosure: Page 9; 80pp; English.
XX
XX This is the modified nucleotide sequence of MUC18 with primer binding
XX sites (AAZ20935 and AAZ20936) to introduce a BamHI site just upstream of
XX the translation start site in order to facilitate cloning.
XX The presence of this glycoprotein has been correlated with the ability
XX of melanomas to metastasize. MUC18 is also associated with normal
XX vascular tissue, and on the smooth muscle of venules, and it expresses
XX sporadically on capillary epithelium.
XX The method can be used as a diagnostic test for prostate cancer which
XX has a relatively high potential for metastasis or which has
```

```
CC metastasized. This allows the physician to choose appropriate surgical,
CC chemo- therapeutic or radiation treatment regimens. In addition,
CC antibody specific to MUC18 can be used to prevent metastasis of prostate
CC cancer cells.
XX
XX Sequence 1962 BP; 457 A; 559 C; 595 G; 351 T; 0 other;
SO
alignment_scores:
Quality: 3316.00 Length: 645
Ratio: 5.157 Gaps: 0
Percent Similarity: 99.690 Percent Identity: 98.760
alignment_block:
US-09-653-961-2 x AAZ20932 ..
Align seq 1/1 to: AAZ20932 from: 1 to: 1962
1 MetGlyLeuProArgLeuValCysAlaPheLeuAlaIaCysCysC 17
|||||
10 ATGGGCTTCCAGGCTGTGTGCGCTTCTTGTCTGCGGCTGCTGCTG 59
|||||
17 SCysProArgValAlaGlyValProGlyCuaIaGluGlnProAlaProG 34
|||||
60 CTGTCTCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 109
|||||
34 IuLeuValGluValGluValGlySerThrAlaLeuLeuLysCysGlyLeu 50
|||||
110 AGCTGTGAGGTGGAAGTGGGAGCAGACCCCTTCTGAACTGGGCTC 159
|||||
51 SerGlnSerGlnGlyAsnLeuSerHisValAspTrpPheSerValHisL 67
|||||
160 TCCCACTCCCAAGGCAACCTCAGCATGTGACTGGTTTCTGTCCACA 209
|||||
67 SGIuYsArgThrLeuIlePheArgValArgGlnGlnGlnGlnSerG 84
|||||
210 GAGAGAGCGAGGCTCATCTTCCGTGTGCGGCGAGGCGAGGCGAGG 259
|||||
84 IuProGlyGluTyrGluGlnArgLeuSerLeuGlnAspArgGlyAlaThr 100
|||||
260 AACCTGGGAGAGACGAGCGGCTCAGCTCCAGGAGAGGAGGCTACT 309
|||||
101 LeuAlaLeuThrGlnValThrProGlnAspGluArgIlePheLeuCysG 117
|||||
310 CTGGCCCTGACTCAAGTCACCCCGGAGAGCGAGCGACTTGTGTGCA 359
|||||
117 NGIuYsArgProArgSerGlnGluTyrArgIleGlnLeuArgValTyrL 134
|||||
360 GGGCAAGCGGCTCGGTCCAGAGTACCGCATCCGCTCCGCTTACA 409
|||||
134 yslAlaProGluGluProAsnIleGlnValAsnProLeuGlyTleProVal 150
|||||
410 AACCTCCGAGAGGAGCAACATCCAGGTCAACCCCTGGGCAATCCGTGTG 459
|||||
151 AsnSerLysGluProGluGluValAlaThrCysValGlyArgAsnGlyT 167
|||||
460 AACAGTAAGAGACCTGAGAGGTGCTACTCTGTAGAGAGAGACGGGTA 509
|||||
167 rProIleProGlnValIleTrpTyrLysAsnGlyArgProLeuLysGlu 184
|||||
510 CCCCATCTCTCAAGTCATGTGTACAAAGATGGCGGCTCTGAAGAGG 559
|||||
184 IuLysAsnArgValHisIleGlnSerSerGlnThrValGluSerGly 200
|||||
560 AGAAGAACCGGCTCCATTCATGTCGTCAGACCTGTGAGTCAAGTGTG 609
|||||
201 LeuTyrThrLeuGlnSerIleLeuLysAlaGlnLeuValLysGluAsp 217
|||||
610 TTGTACACCTTCAGAGTATTCGAAAGGCAAGCTGTGTTAAAGAAAGCA 659
|||||
217 sAspAlaGlnPheTyrCysGluLeuAsnTyrArgLeuProSerGlyAsn 234
|||||
660 AGATGCCCGATTTCATCTGTAGCTCAACTACCGGCTGCCAGTGGGAAC 709
|||||
```

```

234 isMetLysGluSerArgGluValThrValProValPheTyrProThrGlu 250
|||||
710 ACATGAAAGAGTCCAGGGAGTGCACCTGCTTTTCTTACCCAGAGAA 759
251 LysValThrPLeuGluValGluProValGlyMetLeuLysGluLysPAr 267
|||||
760 AAAGTGTGCTGGAAGTGGAGCCCTGGGAATGCTGAAGGAAGGGACCG 809
267 gValGluLeuArgCysLeuAlaAspGlyAsnProProProHisSerI 284
|||||
810 CGTGAATATCAGGTGTTGGCTGATGGACCTCCACACACTTCAGCA 859
284 IeSerLysGluAsnProSerThrArgGluAlaGluGluGluThrAsn 300
|||||
860 TCAGCAACAGAACCCCGACAGGGGCGAGAGGAAGACACACACAC 909
301 AspAsnGlyValLeuValLeuValProValArgLysGluHisSerGly 317
|||||
910 GACAAAGGGGTCTGCTGTGTGGAGCTGCCCGGAAGGACACAGTGGCG 959
317 gTyrGluCysGlnGlyLeuAspLeuAspThrMetIleSerLeuLeuSer 334
|||||
960 CTATGAATGTGAGGCTGGAACCTTGAGACCATGATATGCTGCTGAGTG 1009
334 LuProGlnGluLeuLeuValAsnTyrValSerAspValArgValSerPro 350
|||||
1010 AACCAAGAGAACTACTGTGTAACATATGTCAGCTCCGACAGTGAATGCC 1059
351 AlaAlaProGluArgGlnGluGlySerSerLeuThrLeuThrCysGlu 367
|||||
1060 GCAGCCCTTGAGAGACAGGAAGGACAGCCTACCTGACCTGTGAGGC 1109
367 agLysSerSerGlnAspLeuGluPheGlnThrLeuArgGluGluThrGly 384
|||||
1110 AGAGATGAGCCAGACCTCGAGTTCAGTGGCGAGAGAAGAGACGAGC 1159
384 InValLeuGluArgGlyProValLeuGlnLeuHisAspLeuLysArgGlu 400
|||||
1160 AGGTGCTGGAAGGGGGCCCTGCTTCAGTGTGATGATGACCTGAAAGGG 1209
401 AlAGlyGlyGlyTyrArgCysValAlaSerValProSerIleProGlyLe 417
|||||
1210 GCAGGAGGGCGGCTATGCTGCGGTGGTGTGCCAGATACCGGCTT 1259
417 uAsnArgThrGlnLeuValAsnValAlaIlePheGlyProProTyrPMe 434
|||||
1260 GAACCCGACACAGCTGTGCAAGCTGGCCATTTTGGCCCCCTGTGATGG 1309
434 IapheLysGluArgLysValTyrValLysGluAsnMetValLeuAsnLeu 450
|||||
1310 CATTCAGAGAGAGAGAGTGTGGGTGAAGAATATGTGTGTAATCTG 1359
451 SerCysGluAlaSerGlyHisProArgProThrIleSerTyrAsnValAs 467
|||||
1360 TCTTGTGAAGCGTCAGGACCCCGGCGCCACACATCTCTGGAACGTCAA 1409
467 nGlyThrAlaSerGluGlnAspGlnAspProGlnArgValLeuSerThr 484
|||||
1410 CGGACAGGCAAGTGAACAAGACCAAGATCCACAGAGATCCTGAGCACCC 1459
484 euAsnValLeuValThrProGluLeuLeuGluThrGlyValGluCysThr 500
|||||
1460 TGAATGTCTCTGTGACCCCGGAGCTGTGGAGACAGGTGTGAATGCACG 1509
501 AlaSerAsnAspLeuGlyLysAsnThrSerIleLeuPheLeuGluVal 517
|||||
1510 GCCCTCAAGAGACTGGGCAAAAACCGCATCTCTTCCGTGGAGCTGGT 1559
517 IasnLeuThrThrLeuThrProAspSerAsnThrThrGlyLeuSerI 534
|||||
1560 CAATTAAACACCCCTCACACACAGACTCCACACCAACACAGTGGCTCAGCA 1609

```

```

534 InSerThrAlaSerProHisThrArgAlaAsnSerThrSerThrGluArg 550
|||||
1610 CTTCCACAGTCCAGTCTCATACAGAGCCAAACAGACCTCCACAGAGAGA 1659
551 LysLeuProGluProGluSerArgGlyValValIleValAlaValIleVa 567
|||||
1660 AAGCTGCCGGAGCCGGAGAGCCGGGGCGTGTGATCTGCTGTGANTGT 1709
1567 lCysIleLeuValLeuAlaValLeuGlyAlaValLeuTyrPheLeuTyr 584
|||||
1710 GTGCATCTCTGTGCTGTGGGCTGTGGGCTGTCTCTTATTTCTCTTANA 1759
584 yLysGlyLysLeuProCysArgArgSerGlyLysGlnGluIleThrLeu 600
|||||
1760 AGAAGGGCAAGCTCCGTCAGGGCGCTCAAGGAAGCAAGGAGATCACGCTG 1809
601 ProProSerArgLysSerGluLeuValValGluValLysSerAspLysLe 617
|||||
1810 CCCCCTCTCTTAAGACCGAACTTGATGTTGAAGTTAAAGTCAGATTAAGCT 1859
617 uProGlnGluMetGlyLeuLeuGlnGlnGlySerSerGlyAspLysArgAlaP 634
|||||
1860 CCCAGAAAGATGGGCTCTCTGCAAGGCGACACAGCGGTGACAAAGAGGCTC 1909
634 roGlyAspGlnGlyGluLysTyrIleAspLeuArg 645
|||||
1910 CGGAGAGCAAGGAGAGAAATACATCATCTGTGAG 1944
seq_name: /sids5/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AA61554
seq_documentation block:
ID AA61554 standard; cDNA, 524 BP.
XX
XX AA61554;
AC
XX
XX 19-FEB-2001 (first entry)
DT
XX
XX Genetic suppressor elements which is derived from Mel-CAM cDNA.
DE
XX
XX Genetic suppressor element; melanoma; cell adhesion protein; Mel-CAM;
KW beta-integrin; invasive melanoma; retrovirus vector; solid tumour;
KM angiogenesis; ss.
XX
XX Synthetic.
OS Homo sapiens.
XX
XX MO200058501-A2.
PN
XX
XX 05-OCT-2000.
PD
XX
XX 24-MAR-2000; 2000WO-US07807.
PF
XX
XX 26-MAR-1999; 99US-0126479.
PR
XX
XX (WIST-) WISTAR INST.
PA
XX
XX HeLyn M, Salyamoorthy K;
PI
XX
XX WPI; 2000-594646/56.
DR
XX
XX Trans-recoverable packaging deficient retrovirus vectors for the
PT treatment of solid tumors -
XX
XX Claim 19; Fig 3A; 84pp; English.
XX
XX AAC61554-74 represent genetic suppressor elements which exhibit an
CC anti-melanoma effect. The polynucleotides are derived from a portion
CC of cDNA corresponding to the cell adhesion protein Mel-CAM or
CC beta-integrin. The genetic suppressor elements either inhibit expression
CC or induce hyper-expression of Mel-CAM or beta-integrin. Expression of
CC these proteins are known to be correlated with survival and growth of
CC invasive melanomas. The specification describes methods for inhibiting
CC a phenotype associated with diseased cells. The methods use gene

```

CC suppressor elements for inhibiting such phenotypes. The method involves
CC using a trans-recoverable packaging deficient retrovirus vector to
CC deliver a plurality of polynucleotides to a population of target cells.
CC The methods are used for treatment of solid tumours and for inhibiting
CC angiogenesis in solid tumour cells.

Sequence 524 BP; 124 A; 139 C; 166 G; 95 T; 0 other;

[illegible]

alignment_block: US-09-653-961-2 x AAC61554 ..

Align seg 1/1 to: AAC61554 from: 1 to: 524

327	thrmellieserleuileuSerGluProGlnGluLeuLeuValAsnTyrVa	343	1seraspValaArgValserProAlaAlaProGluIarGlnGluGlySerS	360
328		344		361
329		345		362
330		346		363
331		347		364
332		348		365
333		349		366
334		350		367
335		351		368
336		352		369
337		353		370
338		354		371
339		355		372
340		356		373
341		357		374
342		358		375
343		359		376
344		360		377
345		361		378
346		362		379
347		363		380
348		364		381
349		365		382
350		366		383
351		367		384
352		368		385
353		369		386
354		370		387
355		371		388
356		372		389
357		373		390
358		374		391
359		375		392
360		376		393
361		377		394
362		378		395
363		379		396
364		380		397
365		381		398
366		382		399
367		383		400
368		384		401
369		385		402
370		386		403
371		387		404
372		388		405
373		389		406
374		390		407
375		391		408
376		392		409
377		393		410
378		394		411
379		395		412
380		396		413
381		397		414
382		398		415
383		399		416
384		400		417
385		401		418
386		402		419
387		403		420
388		404		421
389		405		422
390		406		423
391		407		424
392		408		425
393		409		426
394		410		427
395		411		428
396		412		429
397		413		430
398		414		431
399		415		432
400		416		433
401		417		434
402		418		435
403		419		436
404		420		437
405		421		438
406		422		439
407		423		44

seq_name: /SIDS5/gcgcdata/geneseq/geneseqn-emb1/NA2000.DAT:ACC61560

seq_documentation_block:

XX
0061E60.

DT 19-FEB-2001 (first entry)

Genetic suppressor elements which is derived from mel-CAM cDNA

XX	Genetic suppressor element; melanoma; cell adhesion protein; Mel-CAM.
KX	beta-integrin; invasive melanoma; retrovirus vector; solid tumour;
KW	angiogenesis, ss.
XX	
XX	Synthetic.
OS	Homo sapiens.
KX	
PN	W0200058501-A2.

XX AAC61524-74 represent genetic suppressor elements which exhibit an anti-melanoma effect. The polynucleotides are derived from a portion of cDNA corresponding to the cell adhesion protein Mel-CAM or beta-integrin. The genetic suppressor elements either inhibit expression of or induce hyper-expression of Mel-CAM or beta-integrin. Expression of these proteins are known to be correlated with survival and growth of invasive melanomas. The specification describes methods for inhibiting a phenotype associated with diseased cells. The methods use gene suppressor elements for inhibiting such phenotypes. The method involves using a trans-recoverable packaging deficient retrovirus vector to deliver a plurality of polynucleotides to a population of target cells. The methods are used for treatment of solid tumours and for inhibiting angiogenesis in solid tumour cells.

aa Sequence 422 BP; 103 A; 125 C; 118 G; 76 T; 0 other;
sq

[illegible]

```
alignment_block:
```

Align seg 1/1 to: AAC61560 from: 1 to: 422

504 AspleugilylYsaantHserIleleubheleugluleuValasnLeuTh 520
1 GACCTGGGCAAAACCCAGCATCTCTCTCTCGAGCTGGTCATTAAAC 50
520 rThleuThrProaspSerAsnThrThrThrGlyLeuSerThSerThra 537
51 CACCTTCACACCGAGCTCCACACAAACACAGCGCTCAGCACTTCCACTG 100
537 laSerProHistiThrArgAlaasnSerThSerThThGluArgYLeuPro 553
101 CCACTCTCTACCAAGAGCAACAGCACTCCACAGAGAGAAAGCTCCG 150
554 GlueProGluSerArgGlyValValIleValAlaValIleValCysIlele 570
151 GAGCGGAGAGCCGGGGCGGGATCGGGCTGTGATTGTGTCAATCT 200
570 uValLeuAlaValLenglyValAlaValLeuYrPheLeuYrIysLysGlyL 587
201 GGCTCTGGCGGTGTGTGGCCCTCTCTCTCATTTCTCTATAAGAAAGGCA 250

587 yslLeuProCysArgArgSerGlyLysGlnGluIleThrLeuProSer 603
 251 AGCTGCGCTGCGAGCGCTCAGAGGAGCAGAGATCAGCTGCCCGCTCT 300
 604 ArgLysSerGluLeuValAlaGluValLysSerAspLysLeuProGlu 620
 301 CGTAAAGACCGAACTTGTAGTTGAAGTTAACTCAGATTAACTCCAGAGA 350
 620 uMetGlyLeuGlnGlnLysSerGlyAspLysArgAlaProGlyAspG 637
 351 GATGGGCTCTCTGCGAGGCGAGCGGTGACAAAGGGCTCGGAGAGCC 400
 637 lnglyGluLysTyrIleAsp 643
 401 AGGAGAGAAATACATCAT 420

seq_name: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AA00314

seq_documentation_block:

ID AAC00314 standard; cDNA; 441 BP.

AC AAC00314;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 312.

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

PA (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI: 2000-500381/45.

DR P-PSDB; AAG00308.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 diagnostic, forensic, gene therapy and chromosome mapping procedures -
 Claim 1; SEQ ID 312; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. An ORF has been identified within the
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
 CC derived from 30 different tissues. EST sequences usually correspond
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
 CC well suited for isolating cDNA sequences derived from the 5' ends of
 CC mRNAs and even in those cases where longer cDNA sequences have been
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
 CC gene therapy and chromosome mapping procedures. They are used to obtain
 CC upstream regulatory sequences and to design expression and secretion
 CC vectors.

SO Sequence 441 BP; 72 A; 149 C; 135 G; 81 T; 4 other;

alignment_scores:

Quality: 682.00 Length: 136
 Ratio: 5.128 Gaps: 1
 Percent Similarity: 97.794 Percent Identity: 97.794

alignment_block:

US-09-653-961-2 x AAC00314 ..

Align seg 1/1 to: AAC00314 from: 1 to: 441

1 MetGlyLeuProArgLeuValCysAlaPheLeuLeuAlaIaCysCysC 17
 34 ARGGGAGCTCCAGAGCTGTGTCGCCCTTCTTGCTCGCGGCTGCTG 83
 17 scYsProArGValAlaGlyValProGlyGluAlaGlnProAlaProG 34
 84 CTGTCTCGCGCGCGGGGTGTCCCGGAGAGGCTGAGCAGCTCGGCTG 133
 34 luleuValGluValGluValGlySerThrAlaLeuLeuLysCysGlyLeu 50
 134 AGCTGTGAGAGTGAAGTGGGACAGACCCCTTCTGAAGTGGGCTC 183
 51 SerGlnSerGlnGlyAsnLeuSerHisValAspTrpPheSerValHisLy 67
 184 TCCAGTCCCAAGGCAACCTCAGCCATGTGACATGTTTCTGTCCAA 233
 67 sgluLysArgThrLeuIlePheArgValArgGlnGlyGlnGlyLys 84
 234 GAGAGAGCGAGCGCTATCTTCGTGCGCAGGCGCAGGCGCAGAGCG 283
 84 luproGlyGluTyrGlnGlnArgLeuSerLeuGlnAspArgGlyAlaThr 100
 284 AACCTGGGAGTACGAGCAGCGGCTCAACCTCCAGAGCAGAGGGGCTACT 333
 101 leuAlaLeuThrGlnValThrProGlnAspGluArgIlePheLeuCys.G 117
 334 CTGGCCCTGACTCAAGTCACCCCAAGACGAGCGCATCTTCTGTGCAN 383
 117 lnglyLysArgProArgSerGlnGluTyrArgIleGlnLeuArgValTyr 133
 384 GTNRMAAGCGCCCTCGTCCGAGAGTACGCGATCCAGCTCGGCTTAC 433
 134 LysAla 135
 434 AAAGCT 439

seq_name: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AA061562

seq_documentation_block:

ID AAC61562 standard; cDNA; 400 BP.

AC AAC61562;

DT 19-FEB-2001 (first entry)

DE Genetic suppressor elements which is derived from Mel-CAM cDNA.

KW Genetic suppressor element; melanoma; cell adhesion protein; Mel-CAM;
 beta-integrin; invasive melanoma; retrovirus vector; solid tumour;

KW angioneiis; ss.

OS Synthetic.

OS Homo sapiens.

PN WO200058501-A2.

PD 05-OCT-2000.

PF 24-MAR-2000; 2000MO-US07807.

PR 26-MAR-1999; 99US-0126479.

PA (WIST-) WISTAR INST.

```
PI Herlyn M. Salyamoorthy K;
XX
DR WPI; 2000-594646/56.
XX
PT Trans-recoverable packaging deficient retrovirus vectors for the
XX treatment of solid tumors -
XX
PS Claim 19; Fig 11B; 84pp; English.
XX
CC AC61554-74 represent genetic suppressor elements which exhibit an
CC anti-melanoma effect. The polynucleotides are derived from a portion
CC of cDNA corresponding to the cell adhesion protein Mel-CAM or
CC beta-Integrin. The genetic suppressor elements either inhibit expression
CC or induce hyper-expression of Mel-CAM or beta-Integrin. Expression of
CC these proteins are known to be correlated with survival and growth of
CC invasive melanomas. The specification describes methods for inhibiting
CC a phenotype associated with diseased cells. The methods use gene
CC suppressor elements for inhibiting such phenotypes. The method involves
CC using a trans-recoverable packaging deficient retrovirus vector to
CC deliver a plurality of polynucleotides to a population of target cells.
CC The methods are used for treatment of solid tumours and for inhibiting
CC angiogenesis in solid tumour cells.
XX
XX Sequence 400 BP; 95 A; 98 C; 133 G; 74 T; 0 other;

Alignment_scores:
Quality: 655.00 Length: 132
Ratio: 5.038 Gaps: 0
Percent Similarity: 98.485 Percent Identity: 95.455

Alignment_block:
US-09-653-961-2 x AAC61562 ..

Align seg 1/1 to: AAC61562 from: 1 to: 400

315 serGIyArTyrgLUcYgInGlyLeuAspLeuAspThrMetIleSerIe 331
|||||
3 AGTGGCGCTATGATGATGTCAAGCGCTGGAACTTGACACCATATATCGCT 52
|||||

331 uLeuSerGIuPProGInGluLeuValAsnTyValSerAspValArgV 348
|||||
53 GCTGAGTGAACACAGACGACTGTGTAATGTCTGACGTCGCGAG 102
|||||

348 aLsePrrroAlaAProGluArgInGluGlySerSerleuThr 364
|||||
103 TGAgtCCCGAGCCCTGTGAAGACAGAGAGGACGACCTCACCCCTGACC 152
|||||

365 CyAGlUAlAGluseSerGIAspLeuGluPheGInTrPLeuArgGluG 381
|||||
153 TGTGAGGACAGAGAGTACCGACGACCTCGAGTTCCAGTGGCTGAGAGAGA 202
|||||

381 uThrGIyGInValLeuGluArgGlyProValLeuGInleuHIsAspLeuL 398
|||||
203 GACAGACACAGGTGCTGAAAGGGGGCGCTGTCTTCAGTTGCATGACCTGA 252
|||||

398 ySArGIuAlAGlYgLYgLYTyrArGcYsValAlaSerValProSerIle 414
|||||
253 AACGGGAGGAGAGAGCGGCTATCGCTGCTGGCGTCTGCGCCACATTA 302
|||||

303 CCGGGCGCTGAACCGCACACAGCTGTGTCAGTGGCATTTTGGCCCCC 352
|||||

415 ProGlyLeuAsnArGThrGInleuValAsnValAlaIlePheGlyProP 431
|||||
431 oTPMeAlaPheLySGluArgYsValTrpValLySGluAsnMet 446
|||||

353 TTGGATGGCATTCAGAGAGAGAGAGGTGTGGTGAAGAGCAATATG 398
|||||

seq_name: /std5/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:ABA83083
seq_documentation_block:
ID ABA83083 standard; DNA; 2402 BP.
XX
```

```
AC ABA83083;
XX
XX 08-FEB-2002 (first entry)
XX
DE Lutheran blood group protein (BCAM) ovarian tumour marker gene, #5.
XX
XX Ovarian tumour marker gene; human; overexpression; upregulation;
XX epithelial tumour; cancer; diagnosis; prognosis; disease monitoring;
XX identification; serous cystadenoma; borderline serous tumour;
XX serous cystadenocarcinoma; mucinous cystadenocarcinoma;
XX mucinous cystadenoma; borderline mucinous tumour; endometrioid carcinoma;
XX undifferentiated carcinoma; clear cell adenocarcinoma; cystadenofibroma;
XX adenofibroma; Brenner tumour; serial analysis of gene expression; SAGE;
XX immune response pathway; cell proliferation regulation; protein folding;
XX membrane localised; secreted; therapeutic target; cytostatic;
XX gene therapy; vaccine; ds.
XX
XX Homo sapiens.
XX
XX WO200175177-A2.
XX
XX 11-OCT-2001.
XX
XX 03-APR-2001: 2001WO-US10947.
XX
XX 03-APR-2000: 2000US-194336P.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Morin PJ, Sherman-Baust CA, Pizer ES, Hough CD;
XX P-PsDB; ABB50259.
XX
XX WPI: 2001-626450/72.
XX
XX Detecting and identifying ovarian tumor, identifying increased risk for
XX developing ovarian cancer, and determining effectiveness of ovarian
XX cancer treatment, by measuring expression level of ovarian tumor marker
XX gene -
XX
XX Claim 22; page 54; 140pp; English.
XX
XX The invention relates to methods for diagnosing and prognosing ovarian
XX tumors in an individual via the detection and measurement of the
XX expression of ovarian tumor marker genes (ABA83081-ABA83122, ABA83180,
XX ABA83182 and ABA83183) or segments thereof (ABA83123-ABA83169, ABA83179,
XX ABA83181 and ABA83183). The methods of the invention are useful for
XX detecting an ovarian tumor in a patient, for identifying an individual
XX at increased risk for developing ovarian cancer, in prognostic tests for
XX assessing the relative severity of ovarian cancer, in tests for
XX monitoring a patient in remission from ovarian cancer and in tests for
XX monitoring disease status in a patient being treated for ovarian cancer.
XX The methods can additionally be used to identify a particular tumor as
XX being an ovarian tumor (i.e., an epithelial ovarian tumor selected from
XX serous cystadenoma, borderline serous tumour, serous cystadenocarcinoma,
XX mucinous cystadenoma, borderline mucinous tumour, mucinous
XX cystadenocarcinoma, endometrioid carcinoma, undifferentiated carcinoma,
XX clear cell adenocarcinoma, cystadenofibroma, adenofibroma and Brenner
XX tumour. The ovarian tumour marker genes of the invention were identified
XX using SAGE (serial analysis of gene expression) and were found to be
XX overexpressed in a broad variety of ovarian epithelial tumour cells
XX relative to normal ovarian epithelial cells. The marker genes are
XX implicated in immune response pathways, in the regulation of cell
XX proliferation and in protein folding, and many of these are membrane-
XX localised or secreted. In addition to their use as diagnostic and
XX prognostic markers, the ovarian tumour marker genes or their encoded
XX proteins may be used as therapeutic targets for the treatment and
XX prevention of ovarian cancer. Sequences ABA83081-ABA83122, ABA83180,
XX ABA83182 and ABA83184 represent the ovarian tumour marker genes of
XX the invention.
XX
XX Sequence 2402 BP; 426 A; 844 C; 742 G; 390 T; 0 other;
XX
```


326TACGACTGACTCCAGGGGCGC.....CTGTGCT 357
103 uThrGlnValThrProGlnAspGluArgIlePheLeuCys.....GlnG 118
118 LysArgProArgSerGlnGluTyrArgIleGlnLeuArgValTyrLys 134
408 CAGGGGCGGAGCACTGTCAGGCGCACTGGCGGCTCAACAGCTGTTCGA 457
135 AlaProGlnGluProAlaGlnIleGlnValAspProLeuGlyIleProValAs 151
458 AAGCAGAGGCGCACTGAGGTCTCCCAACCAAGGAGCACTGTCTGTAT 507
151 uSerLysGluProGlnGluValAlaThrCysValGlyArgAsnGlyTyrP 168
508 GGAGGACTCTGCCAGAGATCGCAGCTCTCAACAGCCGAGAGGAGACC 557
168 rolleProGlnValIleTyrTyrLysAsnGlyArgProLeuLys..... 182
558 CGGCCCCCAAGATCACGTGTATCGCAACGCGAGCGCTGAGGTGCC 607
183 GlnGluLysAsnArgValHisIleGlnSerSerGlnThrVal...GluSe 198
608 GTACAGATGAACCCAGAGGCGCTACATGACCGACGCGCTCGGAGAGC 657
198 rSerGlyLeuTyrThrLeuGlnSerIleLeuLysAlaGlnLeuValLysG 215
658 CTGGGGCTCTCTCCCTCACGACCTTACCTGCGGCTCCGCAAG 707
215 LysPlysAspAlaGlnPheTyrCysGluLeuAsnTyrArgLeuProSer 231
708 ATGACGAGAGCGCCAGCTTCACGTGCGCGCCACCTACAGCTGCCAG 757
232 GlyAsnHisMetLys...GluSerArgGluValThrValProValPheTyr 247
758 GGCGGCGAGCGCGCTGAGACAGCCCACTTCCACCTCACCCTGACTA 807
247 rProThrGluLysVal.....TyrLeu.....GluValGluProValG 260
808 TCCGACGAGACGACGTGCTGCTGGTGGGAGCCCGCCACCCGACAG 857
260 LysMetLeuLysGlnGluAspArgValGluIleArgCysLeuAlaAspGly 276
858 GCTGGGTACGAGAGGTGACACTGTCCAGCTCTGCGGGGAGAGCGC 907
277 AsnProProHisPheSerIleSerLysGlnAsnProSerThrArgGly 293
908 AGCCCAAGCCGAGATATACGCTTTTCGC.....CTTCAGA 945
293 uAlaGlnGluGluThrThrAsnAspAsn.....GlyValLeuValLeuG 308
946 TGACGAGAGGAGATGTGATGATCTGAGAGGAGAACTGTGACCTGG 995
308 LysProAlaArgLysGlnHisSerGlyArgTyrGluCysGlnGlyLeuAsp 324
996 AAGGAGTACCCGAGGCGGAGCGGAGCTATGCTGACAGAGTGAGAG 1045
325 LeuAspThrMetIleSerLeu...LeuSerGluProGlnGluLeuVal 340
1046 TACGACGGGAGGAGATGACGTCTCCCAAGAGCGCTGAGAGCGCCGT 1095
340 LysTyrValSerAspValArgValSerProAlaAlaProGluArgGln 357
1096 GGCCATATCTGAGCCCGTGGAGCTCAGC.....G 1124
357 LysGly.....SerSerLeuThrLeuThrCys 365
1125 AGGGAGAGGCTTTCTTACTTAAACAGACAGCTGAGCTGAACTGC 1174
366 GluAlaGluSerSerGlnAspLeuGluPheGlnThrLeuArgGluGln 382
1175 TCCGTGACGCGCTGCGCACCCCTGACGTGAGCAAGACTCCAC 1224

382 rGlyGlnValLeuGluArgGlyProValLeuGlnLeuHisAspLeuLysA 399
1225 T.....CCCTGGGCGATGGCCCCATGCTGTCTGATTTCTATACCT 1268
399 rGlnAlaGlyGlyGlyTyrArgCysValAlaSerValProSerIlePro 415
1269 TCGATTCGAATGGCACTACGTATGTGAGGCTCCCTGCCACAGTCCCG 1318
416 GlyLeuAsnArgThrGlnLeuValAlaIlePheGlyProProThr 432
1319 GTCTTCAGCCGACCCAGAACTTCACTGCTGCGCGGCGCATCAGACCC 1368
432 PheAlaPheLysGluArgLys.....ValTyrPValLysG 444
1369 GCTAAAGACAGCGAATATAGCCCAAGCAGATGACACTGG...AGG 1415
444 LysAsnMetValLeuAsnLeuSerCysGlnAlaSerGlyHisProArgPro 460
1416 AAGGAGACGAAGTCACACTCATCTGCTGCGCGGCGCATCAGACCC 1465
461 ThrIleSerThrPasnValAsnGlyThrAlaSerGlnGlnAspGlnAspPr 477
1466 AAACCTGACGTGAGAGCAATTTGGG...GGCAGCCCCGAGAGCAATCC 1512
477 oGlnArg.....ValLeuSerThrLeuAsnValLeuValThrProG 491
1513 CGAGCGCAGAGGTTGGGTGAGCAGCTCTGACCTGAAAGTACACAGC 1562
491 LeuLeuGlnGluThrGlyValGluCysThrAlaSerAsnAspLeuLys 507
1563 CCTGTAGCCGCGATGAGCATCTCTGTGAGACCTTCAACCCGAGGAGAC 1612
508 AsnThrSerIleLeuPheLeuGlnLeuValAsnLeuThrThrLeuThr 524
1613 AAGCGCAGTCTTCCACTTCCGCGCGGAGC..... 1645
524 oAspSerAsnThrThrGlyLeuSerThrThrAlaSerProHisT 541
1645 1645
541 hrArgAlaAsnSerThrSerThrGluArgLysLeuProGluProGluSer 557
1646CCCAAGACTCCAG 1660
558 ArgGlyValValIleValAlaValIleValCysIleLeuValLeuAlaVal 574
1661 GCTGAGAGTGGCCGTATGCGCTGCGCGCTGAGCTGCTGCTGCTCT 1710
574 LLeuGlyAlaValLeuTyrPheLeuTyrLysGlyLysGlyLeuProCysA 591
1711 CGTCTGCTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1757
591 rArgSerGlyLysGlnGluIleThrLeuProProSerArgLysSerGlu 607
1758 GCCGACGAGCGGCGGAGAGAGGCGCTCCGCCCA.....GGGAG 1798
608 LeuValAlaGluValLysSerAspLysLeuProGlnGluMetLysLeu 624
1799 CAGGGGCTGAGACCACTCGGGGTCTGGAGCAACAGAGAGCCGCTCT 1848
624 uGlnGly.....SerSerGlyAspLysArgAlaProGlyAspGlnGlyG 639
1849 CATGGAGAGTCCCTCGAGAGCAAGAGGTGGCAGCGGGGCTTGGAG 1898
639 LysLys 640
1899 ACGAG 1903

seq.name: /sids5/gcycdata/geneseq/geneseq-emb1/NA2001B.DAT:AA572009
seq_documentation_block:
ID AA572009 standard; cDNA; 2402 bp.

AC AAS72009;
 XX 13-FEB-2002 (first entry)
 DT
 DE DNA encoding novel human diagnostic protein #7813.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 PD
 XX 11-OCT-2001.
 PF
 XX 30-MAR-2001; 2001WO-US08631.
 PR
 XX 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 PA
 XX (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI: 2001-639362/73.
 DR P-PSDB; ABG07822.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 1; SEQ ID NO 7813; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent novel human
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcr_sequences.
 XX
 XX Sequence 2402 BP; 390 A; 742 C; 844 G; 426 T; 0 other;
 XX

[illegible]

```

293  ualgluulvulthThrAlaAspAsp.....GlyValLeuValLeuG
294  ::::|||||:::  ||  ||  ||  ||:::|||||
1457  TGAGAGAGAGAGAGTGTGTAATGTGATCTCGAGGGAACTTGACCTGG
308  LuProAlaValGlyGluHisSerGlyArgGlyGlyGlyGlyLeuAsp
309  ::::|||||:::  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
1407  AGGAGATGACCAGGAGCCAGAGGAGGAGGACCTATGGCTCGACAGTGAAGAT
325  LeuAspThrMetLeuSerLeu...LeuSerGluProGlnLeuLeuVal
326  ::::|||||:::  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
1357  TACGACGGGCGAGATACGTGGAGGTCTCCAAAGACGCTGAGGTGAGCGGT
340  AsnIleValSerAspValArgValSerProAlaAlaProGluAlaGlnG
341  ::::|||||:::  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
1307  GGCCATCTGTGAGCCCTCGAGCTGACG.....G
357  LuGly.....SerSerLeuThrLeuIleProGly
358  ::::|||||:::  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
1278  AGGGAGAGGTCTTTCTCTTACCTCTAAACACACATGTGACAGTGTGACCTGC
366  GluAlaGluSerSerGlnAspLeuPheLeuMetIleProLeuArgGluIle
367  ::::|||||:::  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
1228  TCCGGGACGGCTCCGCCACCCCTGCTACGCTGAGCAAGAGACTCCAC
382  LeuValLeuLeuArgGlyProValLeuGlnLeuHisAspLeuValAsp
383  ::::|||||:::  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
1178  T.....CCCTGGGCGAGATGGCCCATCTGCTGCTGCTGATCTTATACCT
399  TGGUAlaGlyGlyGlyArgArgCysValAlaSerValProSerIlePro
400  ::::|||||:::  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
1134  TCGATTCGAAAGGACCTGATGATGAGGCTCCGCGCCACAGATCCGG
416  GlyLeuAsnArgThrGlnLeuValAsnValAlaIlePheLeuProPro
417  ::::|||||:::  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
1084  GTCCGACGCGGACACCAAGAACTTCAAGCTGCTGATCCAAAGCTGGCACA
432  PheLeuPheLeuGlyGluArgGlyLys.....ValIleProValGlyG
433  ::::|||||:::  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
1034  GTTAAAGACAGCGGAAATATAGAGCCCAAGGACAGATGACAGCTGG...AGGG
444  LuAsnMetValLeuAsnLeuSerCysGluAlaSerGlyAlaIleProArgPro
445  ::::|||||:::  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
987  AAGGAGAGGAAGTACACATCATCTGTCTGCGCCGAGCAATCCAGACGCC
461  ThrIleSerTrpAsnValAlaGlnGlyThrIleAserGlnGlnAspIleAsp
462  ::::|||||:::  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
937  AATCACTACCTGAGACCAATTGGGG...GGCAGCCCGACCAAGCCAAATCC
477  GlnArg.....ValLeuSerThrLeuAsnValLeuValThrProG
478  ::::|||||:::  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
890  CGGACGCGAGGTTGGGTGAGGACGCTCTGACCTCGAAATGATCAGCAGG
491  LuLeuLeuGlnThrGlyAlaGlyGlyThrAlaSerAsnAspLeuGlyLys
492  ::::|||||:::  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
840  CCTTACGCGGAGTGAATCTCTGTGAGGCTCGAACCCGACGGGAAAC
508  AsnThrSerIleLeuPheLeuGluLeuValAlaLeuThrThrLeuThrPr
509  ::::|||||:::  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
790  AAGCCCATGTCTTCCACTCTGGGCGCGCTGAGC.....G
524  AspSerAsnThrThrThrGlyLeuSerThrSerThrAlaSerProIle
525  ::::|||||:::  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
758  .....G
541  hrArgAlaAsnSerThrSerThrGluArgLysLeuProGluProGluSer
542  ::::|||||:::  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
757  .....CCGACAGCTCCGAG
558  ArgGlyValValIleValAlaValIleValCysIleLeuValLeuAlaVal
559  ::::|||||:::  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
742  GCTGGAGAGGCGGCAAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT
574  LeuGlyAlaValLeuThrThrPheLeuThrLysGlyGlyValLeuProCysA

```

[illegible]

alignment_scores:

Quality: 633.00 Length: 123
Ratio: 5.189 Gaps: 0
Percent Similarity: 99.187 Percent Identity: 97.561

alignment_block:
US-09-653-961-2 x AAC61555 ..

Align seg 1/1 to: AAC61555 from: 1 to: 373

```

378 LeuArgGluGluThrGlyValLeuGluArgGlyProValLeuGlnLe 394
|||||
3 CTGAGAGAAAGACAGACACCGCTGAAAGGCGCTGCTCTCAGATT 52
|||||
394 uHisApeLeuysArgGluAlaGlyGlyTyrArgCysValAlaSerV 411
|||||
53 GCATGACCTGAAACGGAGCAGACGCGCTATCCGCTGCGGCTCG 102
|||||
411 aProSerIleProGlyLeuAsnArgThrGlnLeuValAsnValAlaIle 427
|||||
103 TGGCCAGCATACCGGCTGAAACGACACAGCTGCTCAAGCTGCCCAT 152
|||||
428 PheGlyProProTPrMetAlaPheGlyGluArgLysValTTPValIysG 444
|||||
153 TTTGGGCCCCCTTGATGTCATCAAGAGAGAGAGGTGGGTGAAGA 202
|||||
444 uAsnMetValLeuAsnLeuSerCysGluAlaSerGlyHisProArgProT 461
|||||
203 GAATATAGGTGTTGAATCTGTGTGAAGCGTGAAGGACCCCGCCCA 252
|||||
461 hTlleSerTrpAsnValAsnGlyThrAlaSerGluGlnAspGlnAspPro 477
|||||
253 CCATCTCTCGAAGCAGCAACGCGCAGCAAGTGAACAGCCAGATCCA 302
|||||
478 GlnArgValLeuSerThrLeuAsnValLeuValThrProGluLeuGln 494
|||||
303 CACGAGTCTCTGACGACCCCTCAATGTCTCTGACCCCGAGCTGTGGA 352
|||||
494 uThrGlyValGluCysThr 500
|||||
353 GACAGGTGTGAATGCACG 371
|||||
seq_name: /SID55/gcdata/geneseq/geneseq-n-emb1/NA001A.DAT:ABA09243
seq_documentation_block:
ID ABA09243 standard; cDNA; 2617 BP.
AC ABA09243;
XX
XX 11-JAN-2002 (first entry)
XX
XX Human Lutheran blood group glycoprotein homologue cDNA, SEQ:1019.
XX
XX Human; cytokine; cell proliferation; cell differentiation; growth factor;
XX haematopoietic regulation; tissue growth; immunomodulator; activin;
XX inhibin; chemotaxis; chemokinesis; thrombolytic; oncogenesis;
XX proliferation; metastasis; cancer; tumour; haematopoietic disorder;
XX myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
XX chronic inflammatory condition; proliferative retinopathy;
XX atherosclerosis; coronary heart disease; arterial ischaemia;
XX bone disorder; osteoporosis; vascular growth disorder;
XX tissue regeneration; wound healing; infection; immune disorder;
XX cell culture; drug screening; gene therapy; antiinflammatory;
XX antilastmatic; antiallergic; haemostatic; antidiabetic;
XX cytosolic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
XX antifungal; vulnery; antidiacer; ss.
XX
XX Homo sapiens.
XX OS
XX WO200157188-A2.
XX PN
XX 09-AUG-2001.
XX PD

```

```

PF 05-FEB-2001; 2001WO-US03800.
XX
XX 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX MPI: 2001-457740/49.
XX
XX P-PSDB; ABB11999.
XX
XX Human proteins and DNA encoding sequences useful for preventing,
PT treating or ameliorating a medical condition in a mammalian subject
PS e.g. arthritis and cancer.
XX
XX Claim 1: Page 859-860; 1963pp; English.
XX
XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC invention also relates to vectors and recombinant host cells comprising a
CC nucleotide of the invention, methods of producing the novel polypeptides,
CC antibodies against the polypeptides, methods of detecting the nucleotides
CC or polypeptides in a sample, and methods of identifying compounds which
CC bind to polypeptides of the invention. Although novel, many of the
CC polypeptides of the invention have homology to known proteins, and hence
CC giving an insight into their probable biological activities, and hence
CC potential therapeutic applications. The polypeptides of the invention may
CC have various activities, including cytokine, cell proliferation or cell
CC differentiation activities; stem cell growth factor activity;
CC haematopoietic regulatory activity; tissue growth activity;
CC immunomodulatory activity; activin- or inhibin-related activities;
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC thrombolytic activities; receptor or ligand activities; or may be
CC involved in oncogenesis, cancer cell proliferation or metastasis.
CC Depending on their biological activities, polypeptides and nucleotides of
CC the invention are useful for preventing, treating or ameliorating medical
CC conditions, e.g., by protein or gene therapy. Such conditions include
CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC proliferative retinopathy, atherosclerosis, coronary heart disease,
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC vascular growth. Polypeptides involved with tissue regeneration and
CC repair (e.g., of burns, incisions and ulcers), while those with
CC healing (e.g., of burns, incisions and ulcers), while those with
CC immunomodulatory activities may be used in the treatment of viral,
CC bacterial and fungal infections in addition to immune disorders.
CC Polypeptides with growth factor activity may be used in cell cultures to
CC promote cell growth. For example, such polypeptides may be used to
CC manipulate stem cells in culture to give rise to neuroepithelial cells
CC that can be used to augment or replace cells damaged by illness.
CC Autoimmune disease or accidental damage. The polypeptides and nucleotides
CC may also be used in the diagnosis of the above conditions, and in drug
CC screening techniques. The present sequence represents a cDNA encoding a
CC novel human polypeptide of the invention.
XX
XX Sequence 2617 BP; 533 A; 1018 C; 626 G; 439 T; 1 other:

```

alignment_scores:

Quality: 606.00 Length: 640
Ratio: 1.717 Gaps: 26
Percent Similarity: 55.156 Percent Identity: 29.844

alignment_block:
US-09-653-961-2 x ABA09243 ..

Align seg 1/1 to: ABA09243 from: 1 to: 2617

```

2 GlyLeuProArg...LeuValCysAlaPheLeuLeuAlaCysCysCy 17
||| ||||| ||||| ||| |||||
56 GGGGCCCGCGCTGTCTGCTGCGAGTCTGCTGCGCGGCGG..... 97

```

```
17 sCysProArGValAlaGlyValProGlyGluAlaGluGlnProAlaProG 34
98 .....CACCCAGATGCCAGGCGAGGTGGCTTGTCTGTACCCC 137
34 LuLeuValGluValGluValSerThrAlaLeuLeuLysCysGlyLeu 50
138 CGCTGGTGGAGGTGATGCGAGGAAAGTGTGATTCGTGGACTGC..... 181
51 SerGlnSerGlnLysLysLeuSerHis.....ValAspTrpPhe..... 63
182 ...ACCCCTACGGAGACCCACGACCATTAATGCTGGAAATGTTCCCTTAC 228
64 .....SerValHisLysGluLysA 70
229 CGACCGCTCGGAGACTCGCCCGCTAGCTCGGTGAGATGCAAGGCT 278
70 rGThrLeuIlePheArgValArgGlnGlnGlnGlnSerGlnProGly 86
279 CTGAGCTTCAGAGTCAATGACAGACACCGGGGCGGAGTCCGCCA... 325
87 GluTyrGluArgLeuSerLeuGlnAspArgGlyAlaThrLeuAlaLe 103
326 .....TACCAGTGGACTCCACAGGGCGC.....CTGGTCT 357
103 uThrGlnValThrProGlnAspGluArgIlePheLeuCys.....GlnG 118
358 GGGTGAAGGCCAGGTGGCGACGACGAGCTGTGCTGCTGAGG 407
118 LysAspArgSerGlnGluTyrArgIleGlnLeuArgValTyrLys 134
408 CAGGGGGGAGGAGACTGCTGAGGCGGCTCGGGCTCAAGTGTTCGA 457
135 AlaProGluGluProAsnIleGlnValAsnProLeuGlyTleProValAs 151
458 AACGCAAGGCCACTGAGGTCTCCCCCAACAAAGGAGCACTGTCTGTAT 507
151 nSerLysGluProGluGluValAlaThrCysValGlyArgAsnGlyTyrP 168
508 GAGAGACTCTGCCAGAGATCGCCACCAACAGACACCGGAAAGCAACC 557
168 rGLeuProGluValIleTyrTyrLysAsnGlyArgProLeuLys..... 182
558 CGGCCCCCAAGATCACTGATCGCAAGGGGAGCGCTGAGGTGCC 607
183 GluGluLysAsnArgValHisIleGlnSerGlnThrVal...GluSe 198
608 GTAGAGATGAACCAAGAGGCTACATGACCGCAGCGTCCGGAGGCG 657
198 rSerGlyLeuTyrThrLeuGlnSerIleLeuLysAlaGlnLeuValLysG 215
658 CTGGGGCTGTCTCTCCCTACACGACCTCTACCTCGGCTCCGCAAG 707
215 LuAspLysAspAlaGlnPheTyrCysGluLeuAsnTyrArgLeuProSer 231
708 ATGACGAGAGCCAGCTTCACCTGCGCGCCACTACAGCTCCGCGGAG 757
232 GlyAsnHisMetLys...GluSerArgGluValThrValProValPheTy 247
758 GGGCGGCAAGGCGGCTGACAGCCCACTTCACCTCACCTGCCTCACTA 807
247 rProThrGluLysVal.....TrpLeu.....GluValGluProValG 260
808 TCSCAGGAGCAGTGCAGACTCTGGGTGGGCGCCGCTCCACCCAGCAG 857
260 LysMetLeuLysGlnGluLysArgValGluIleArgCysLeuAlaAspGly 276
858 GTGGGTACGCGAGGTGACACTGTCCAGCTGCTCCGGGGGAGGAGCG 907
277 AsnProProHisPheSerIleSerLysGlnAsnProSerThrArgG 293
908 AACCCGAGCCGAGATATACGCTTTCCG.....CTTCAGGA 945
293 uAlaGluGluGluTyrThrAsnAspAsn.....GlyValLeuValLeuG 308
946 TGAGCAGGAGAGAGTCTGTAATGTGAATCTCGAGGGAGACTTGACCTGG 995
308 LuProAlaArgLysGluHisSerGlyArgTyrGluGlnGlnLeuAsp 324
996 AGGAGATGAGACCCGGGCGCAGAGGGGAGACTATGCTGCAAGATGAGGAT 1045
325 LeuAspThrMetIleSerLeu...LeuSerGluProGlnGluLeuValA 340
1046 TACGAGCGCGCAGATGACTGACGCTCTCCAGAGCGCTGAGACTCGGCT 1095
340 LysTyrValSerAspValArgValSerProAlaAlaProGluArgGln 357
1096 GGCCTATCTGGACCCCTGGAGCTCAGC.....G 1124
357 LuGlySerSerLeuThrLeu.....ThrCys 365
1125 AGGGGAAGGTGCTTTCCTTACCTTAACACGAGGCGAGTGTGAATTC 1174
366 GluAlaGluSerSerGlnAspLeuGluPheGlnTyrPheArgGluGluTh 382
1175 TCCGTGACAGGCTGCCACCCCTGCTACGCTGAGACCAAGAGACTCCAC 1224
382 rGlyGlnValLeuGluArgGlyProValProValLeuHisAspLeuLysA 399
1225 T.....CCCTGGGCGATGGCCCATGCTGTGCTCACTTATCACCT 1268
399 rGluAlaGlyGlyGlyTyrArgCysValAlaSerValProSerIlePro 415
1269 TCGATTCCATGGCAGCTACATGATGAGGCTGCTCCCTCCACAGTCCG 1318
416 GlyLeuAsnArgThrGlnLeuValAsnValAlaIlePheGlyProPTr 432
1319 GTCTCAGCGCCGCCAAGAACTTCACTGCTGTGTCGAAGGCTGCCAGA 1368
432 PheAlaPheLysGluArgLys.....ValTyrValLysG 444
1369 GCTAAAGACAGCGGAATAGAGCCCAAGCGAGTGGCAGCTGG...AGG 1415
444 LuAsnMetValLeuAsnLeuSerCysGluAlaSerGlyHisProArgPro 460
1416 AAGGAGCAGAGTCACTCATCTGCTGCTGCCCGCCATCCAGACAGC 1465
461 ThrIleSerTrpAsnValAsnGlyThrAlaSerGlnGlnAspGlnAspPr 477
1466 AAACCTACACTGGAGCAATTGGG...GGCAGCCCGCAGAGCCAAATCC 1512
477 GlnArg.....ValLeuSerThrLeuAsnValLeuValThrProG 491
1513 CGAGCGGAGGAGGTGGGTGAGCAGCTCTGTGACCTGAAAGTGAACAGCG 1562
491 LuLeuLeuGluThrGlyValGlnCysThrAlaSerAsnAspLeuGlyLys 507
1563 CCTGAGCGCGGATGGCATCTCTGTGAAGCTTCAAC..... 1600
508 AsnThrSerIleLeuPheLeuGluLeuValAsnLeuThrThrLeuThrP 524
1601 .....CC 1602
524 AspSerAsnThrThrThrGluLeuSerThrSerThrAlaSerProHisT 541
1603 CCAGGAGAACAAAGCCCATGTCTTCCACTTCCGACACCGTAGCCCCAG 1652
541 hrArgAlaAsnSerThrSerThrGluArgLysLeuProGluProGluSer 557
1653 CC.....TCCAGG 1660
558 ArgGlyValValIleValAlaValIleValCysIleLeuValLeuAlaVa 574
1661 GCTGGAGTGGCGGTACATGCGCTGAGCGGTGAGGCTCTGCTGCTCT 1710
574 IleuGlyAlaValLeuTyrPheLeuTyrIrisLysGlyLysLeuProCysA 591
```

1711 CGTCGTTGCTGCTTCTACTGCTGAGACGCAAGGGGCG...CCCTGCT 1757
591 rgatgserglylysglnglu 597
1758 GCCGCCAGCGCGGGAGAG 1777

seq_name: /STD55/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AA61559

seq_documentation_block:

ID AAC61559 standard; cDNA: 297 BP.

AC AAC61559;

DT 19-FEB-2001 (first entry)

DE Genetic suppressor elements which is derived from Mel-CAM cDNA.

XX Genetic suppressor element; melanoma; cell adhesion protein; Mel-CAM;

KW beta-integrin; invasive melanoma; retrovirus vector; solid tumour;

KW angiogenesis; ss.

XX Synthetic.

OS Homo sapiens.

PN W0200058501-A2.

PD 05-OCT-2000.

PE 24-MAR-2000; 2000WO-US07807.

PR 26-MAR-1999; 99US-0126479.

PA (WIST-) WISTAR INST.

PI Herlyn M, Salyamoorthy K;

XX WPI; 2000-594646/56.

PT Trans-recoverable packaging deficient retrovirus vectors for the

PS treatment of solid tumors -

PS Claim 19; Fig 10B; 84pp; English.

XX AAC61554-74 represent genetic suppressor elements which exhibit an
CC anti-melanoma effect. The polynucleotides are derived from a portion
CC of cDNA corresponding to the cell adhesion protein Mel-CAM or
CC beta-integrin. The genetic suppressor elements either inhibit expression
CC of or induce hyper-expression of Mel-CAM or beta-integrin. Expression of
CC these proteins are known to be correlated with survival and growth of
CC invasive melanomas. The specification describes methods for inhibiting
CC a phenotype associated with diseased cells. The methods use gene
CC suppressor elements for inhibiting such phenotypes. The method involves
CC using a trans-recoverable packaging deficient retrovirus vector to
CC deliver a plurality of polynucleotides to a population of target cells.
CC The methods are used for treatment of solid tumours and for inhibiting
CC angiogenesis in solid tumour cells.

XX Sequence 297 BP; 69 A; 69 C; 103 G; 56 T; 0 other;

alignment_scores: Quality: 501.00 Length: 98

Ratio: 5.165 Gaps: 0

Percent Similarity: 98.980 Percent Identity: 96.939

alignment_block:

US-09-653-961-2 x AAC61559 ..

Align seg 1/1 to: AAC61559 from: 1 to: 297

356 Gllglnglyserleuthrleuthrcysgluaglsuserseqlnas 372
3 CAGGAGGACGACACCTCCTGACCTGAGACGACAGAGTAGCCACAGA 52

372 pleugluhpheglntprleuarglgluflrhlrglgluvalleugluarg 389
53 CCHGAGCTCCAGTGGCTGAGAGACAGACACACGAGCTGGAAAGG 102
389 lProValleuGlnLeuHisAspLeuLysArgGluIaGlyGlyTyr 405
103 GGCCTGGCTTCAGTTGCATGACCTGAACGAGGAGGACGAGGCGCTAT 152
406 ArgCysValAlaSerValProSerIleProGlyLeuAsnArgThrGln 422
153 CGTCGCTGGGCTGTGCCCCACATACCGGCTGACCCGACACAGCT 202
422 uValAsnValAlaIlePheGlyProProTrpMetAlaPheLysGluArgL 439
203 GGTCAAGCTGGCCATTTTGGCCCCCTTGATGGCATTCAGAGAGAGA 252
439 ySValTTPValLysGluAsnMetValLeuAsnLeuSerCysGlu 453
253 AGTGTGGGTGAAGAGATAATGTTGAATCTGCTTGTGA 296

seq_name: /STD55/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAF68399

seq_documentation_block:

ID AAF68399 standard; cDNA: 4235 BP.

AC AAF68399;

DT 12-APR-2001 (first entry)

DE Human lung tumour protein related nucleotide sequence SEQ ID NO:317.

XX Human; lung cancer; lung tumour; lung tumour protein; gene therapy;

KW human cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;

KW cytostatic; antisense inhibition; ss.

XX Homo sapiens.

OS W0200100828-A2.

PN 04-JAN-2001.

PD 30-JUN-2000; 2000WO-US18061.

PE 30-JUN-1999; 99US-0346492.

PR 15-OCT-1999; 99US-0419386.

PR 17-DEC-1999; 99US-0468687.

PR 30-DEC-1999; 99US-0476300.

PR 06-MAR-2000; 2000US-0519642.

PR 22-MAR-2000; 2000US-0533077.

PR 10-APR-2000; 2000US-0546259.

PR 27-APR-2000; 2000US-0560406.

PR 05-JUN-2000; 2000US-0589184.

XX (CORI-) CORIXA CORP.

XX Wang T, Banpur CS, Iodes MJ, Fanger GR, Vedlack TS, Carter D;

XX Retter MW, Mannion J;

XX WPI; 2001-071488/08.

XX Lung tumor-associated proteins and the nucleic acids that encode them,
XX useful for preventing, diagnosing and treating lung cancer -
XX Example 1; Page 244-245; 436pp; English.

87	GLTPTGSLGlnIArgLeuSerLeuGlnAspArgValAlaThrLeuAlaLeu	103
316	GAATACAAAGACAGCATTTGAACCTCTCAGAA...AACTACACTTTGTCTAT	362
103	uThrgInValThrProGlnAspGluArgTllePheLeuGln...GlyL	119
363	CAGTATTCGAAGAGCATCAGTACGATGAAGAAAGATTTGCTCATGCTAATGA	412
119	ysArgProAlaGserGlnGluuYArgTlleGlnLeuArgValTyrLysAla	135
413	CTGAGGACACGCTGTTTGAGGACCTACATATGTCAAGGTGTTCAGCA	462
136	ProGlnGluProAsnIleGlnValAsnProLeuGlyLleProValAsn	152
463	CCATCTAAACCTGAATTTGTAAGCAAGACAGCTTTCTC...GAAC	506
152	rlYsGluProGlnGluValAlaAlaThrCysValGlyAlaGlnGlyTyrProI	169
507	AGGACAGCTAAAAAAGTTGGTGACCTCATTCTCAGAAAGCACTTATCCAG	556
169	leProAlnValIleTyrPylTysAsnGlyArgProLeuGlnGluLys	185
557	ATGGCATATTCATGATGTCACAGAAATGGAAAGTGTACATCCCTGCA	606
186	AsnArgValHisIleGlnSerSerGlnThrValGluSer...SerGlyLe	201
607	GGAGCGGTGCTCATATTTTAAAAAGAAATGAGACCCAGTGCATCAGCT	656
201	uYrThrLeuGlnSerIleLeuLysAlaGlnLeuValLysGluAspLysA	218
657	CTATCCATGACTTCCACCCCTGGAGTACACAAACCAAGGCTGCATAC	706
218	spAlaGlnPheTyrCysGluLeuAsnTyrArgLeuProSerGlyLysHis	234
707	AATGCATTCATCAGCTGCTGGTGACATATTATGACACATCTGGCCAGAA	756
235	MetLysGluSerArgGluValThrValProValPheTyrProThrGlyL	251
757	ACAATTCAATTCGAAACAGGACAGTATTGATTATTCATCTCAGACAGA	806
251	sValTyrLeuGlnVal...GluProValGlyMetLeuGlnGluLysAspA	267
807	GGGACAAATACAAAGTCTCTCCACCAAAAARTGCCATCAAGAAAGGGATA	856
267	rgValGluIleArgCysLeuAlaAspGlyAsnProProHis...Asp	281
857	ACATCACTCTTAATATCTTAAAGGAATGGCAACCTCCCCACAGAGATTT	906
282	...PheSerIleSerLysLysAsnProSerThrArgGluAlaGlnGluG	297
907	TGTGTTTACTTACCGACAGACGCGGAGGAATAAAGAACTCAAAATCTTA	956
297	uThrThrAsnAspAsnGlyValLeuValLeuGluProAlaArgLysGluH	314
957	CACACTGAGGAT...GTGAGGCGCCATG	982
314	IserGlyArgTyrGluCysGlnGlyLeuAspLeuAspThrMetIleSer	330
983	CACACAGACACATCAAGTCTCCCTCATGACAAAAAAAAGCATTTGATCT	1032
331	LeuLeuSerGluProGlnGluLeuLeuValAsnGlyValSerArgValAla	347
1033	TCACACGCC...ATCACAGTTCACTATTTC...GATTTGTC	1066
347	GValSerProAlaAlaPro...GluArgGlnGluGlySerSerLeuThrL	363
1068	CTTAAACCCAAATGGGAGAGTACTATGACAGATGTTGTATGCCCTTAACCG	1112
363	euThrCysGluAlaGluSerSerGlnAspLeuGluPheGlnThrPheArg	379
1118	TGTCAATGACCAATATCTGCTACCGAGAAATGCAACTGTGTATGATGATAA	1167

```

380 GUGLUthrglyGlnValleuGluArgglyProValleuGlnLeuHisAs 396
      ::::: ||||| ::::: ::::: :::::
1168 GAAAGACATCAGG....CTGCATGTACCCG.....TCATTTCTCAG 1205
      ::::: ::::: ::::: :::::
396 pLeuLysArgGluAlaGlyGlyGlyTyrArgCysValAlaSerValProS 413
      ::::: ::::: ::::: :::::
1206 TCTTCATTTACAGATCGTGAAGATCTGTCCGAACCTGCTCTGCAG 1255
      ::::: ::::: ::::: :::::
413 erlleProGlyLeuAsnArgThrglnLeuValAsnValAlaIlePhegly 429
      ::::: ::::: ::::: :::::
1256 AGGTTAAGAGACTAAAGAAAGAGACTCATGTACTCTCATTTGAGAGGC 1305
      ::::: ::::: ::::: :::::
430 ProProTrpMetAlaPheLysGluArgLysValTrrPValLysGluAsnMe 446
      ::::: ::::: ::::: :::::
1306 AAMCCT.....CAATTAATAATGACAAAGAAACGATCCAGTCGACT 1349
      ::::: ::::: ::::: :::::
446 tValleuAsnLeuSerCysGluAlaSerGlyHisProArgProThrlHis 463
      ::::: ::::: ::::: :::::
1350 ATCTAAACAMTATCTGCCATGTGGAAGGTTTCCAAAGCCAGCCATTTC 1399
      ::::: ::::: ::::: :::::
463 erTrpAsnValAsnGlyThrAlaSer.....GluGlnAspGlnAsp 476
      ::::: ::::: ::::: :::::
1400 ARTGGACAATTACTGCGCAGTGGAGCGCTATAACCAACAGAGCAATCT 1449
      ::::: ::::: ::::: :::::
477 ProGln.....ArgValleuSerThrLeuAsnValleuValThrPr 490
      ::::: ::::: ::::: :::::
1450 CCTTATATTAAATGCGAGTATTACT.....AAATATTATCTTTCCCC 1493
      ::::: ::::: ::::: :::::
490 ogGluLeuGluThrGlyValGlyCysThrAlaSerAsnAspLeuGlyL 507
      ::::: ::::: ::::: :::::
1494 TGAAGAGAAATGTACA...TTAACTTGACACAGCAAGAAACCAACCTG 1540
      ::::: ::::: ::::: :::::
507 yAsnThrSerIleLeuPheLeuGluLeuValAsnLeuThrThrLeuThr 523
      ::::: ::::: ::::: :::::
1541 GAACAGTAAACCTCTGAATGTCTCGCTATAAGTATT..... 1578
      ::::: ::::: ::::: :::::
524 ProAspSerAsnThrThrThrGlyLeuSerThrSerThrAlaSerProH 540
      ::::: ::::: ::::: :::::
1579 CCAGAA.....CCAGAA.....CA 1586
      ::::: ::::: ::::: :::::
540 sThrArgAlaAsnSerThrSerThrGluArgLysLeuProGluProGluS 557
      ::::: ::::: ::::: :::::
1587 CGATGAGGCGACGAGATAGATGAAACAGAA...GAAAGGTGAAATG 1633
      ::::: ::::: ::::: :::::
557 eArArgLysValValIleValAlaValIleValAlaCysIleLeuValAla 573
      ::::: ::::: ::::: :::::
1634 ACCAGGCAAACTAATTGGGGAATGCTGTGCTCCCTGCTGCTGCC 1683
      ::::: ::::: ::::: :::::
574 ValleuGlyAlaValleuTyrPheLeuTyrLysLys..... 585
      ::::: ::::: ::::: :::::
1684 CTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1733
      ::::: ::::: ::::: :::::
586 .....GlyLysLeu.ProCysArgArgSer 593
      ::::: ::::: ::::: :::::
1734 ATCAAAACATGTAAACAGAGCTCGGATATGAGAGAAACAAACAAAGT 1783
      ::::: ::::: ::::: :::::
594 gLyLysGlnGluIleThr...LeuProProSerArgLysSerGluLeuVal 609
      ::::: ::::: ::::: :::::
1784 TAGAAGAAACATCAACAAACAGAGCGTAAGAGAGAAACGTCTCTAGT 1833
      ::::: ::::: ::::: :::::
609 lValGlu.ValLysSer.....AspLysLeuProGluGluMetGlyLeu 623
      ::::: ::::: ::::: :::::
1834 TGTCGACAGATATAAATCATATATAGAACCAATTTGAACGATCAACCTG 1883
      ::::: ::::: ::::: :::::
seq_name: /sgn2_6/ptodata/2/1na/PCTUS_COMB.seq: PCT-US93-00031-22
seq_documentation_block:
; Sequence 22, Application PC/TUS9300031
; GENERAL INFORMATION:
; APPLICANT: Osborn, Laurelee
; APPLICANT: Benjamin, Christopher D.
; TITLE OF INVENTION: ANTIBODIES RECOGNIZING THE FOURTH
; IMMUNOGLOBULIN-LIKE DOMAIN OF VCAM1

```

```

? NUMBER OF SEQUENCES: 23
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Allegretti & Witcoff, Ltd.
? STREET: 10 South Wacker Drive
? CITY: Chicago
? STATE: IL
? COUNTRY: US
? ZIP: 60606
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: IBM PC compatible
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: PCT/US93/00031
? FILING DATE: 19930112
? CLASSIFICATION:
? ATTORNEY/AGENT INFORMATION:
? NAME: McNicholas, Janet M.
? REGISTRATION NUMBER: 32,918
? REFERENCE/DOCKET NUMBER: 92,306-A; D001 CIP PCT
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (312) 715-1000
? TELEFAX: (312) 715-1234
? INFORMATION FOR SEQ ID NO: 22:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1941 base pairs
? TYPE: NUCLEIC ACID
? STRANDEDNESS: double
? TOPOLOGY: linear
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1..1941
? FEATURE:
? NAME/KEY: mat_peptide
? LOCATION: 73..1941
? FEATURE:
? NAME/KEY: sig_peptide
? LOCATION: 1..72
? PCT-US93-00031-22

alignment_scores:
      Quality: 244.50      Length: 683
      Ratio: 0.743      Gaps: 32
Percent Similarity: 48.170      Percent Identity: 21.523

alignment_block:
US-09-653-961-2 x PCT-US93-00031-22 ..

Align seg 1/1 to: PCT-US93-00031-22 from: 1 to: 1941

31 ProAlaProGluLeuValGluValGlySerThrAlaLeuLeu 47
   |||::: |||::: |||::: |||::: |||::: |||:::
91 CTTGACCCCGG...ATTGCTGCTCAGATGTGAGACACATGATGATGAC 137
   |||::: |||::: |||::: |||::: |||::: |||:::
47 sCysGlyLeuSerGlnSerGlnGlyAsnLeuSerHisValAspTrpPheS 64
   |||::: |||::: |||::: |||::: |||::: |||:::
138 ATGTAGTGTCAATGGGCTGTGAATCC.....CAATCTTCTCTGG.... 177
   |||::: |||::: |||::: |||::: |||::: |||:::
64 erValHisLysGlnLysArgThrLeuIlePheArgValArgGlnGln 80
   |||::: |||::: |||::: |||::: |||::: |||:::
178 .....AGAACCAGATAGACAGCCCTCTGAGCGGAAG 210
   |||::: |||::: |||::: |||::: |||::: |||:::
81 GlyInSerGluProGlyGlyTyrGlnGlnArgLeuSerLeuGlnAspAr 97
   |||::: |||::: |||::: |||::: |||::: |||:::
211 GTGAGAGTGAG.....GGGAC 227
   |||::: |||::: |||::: |||::: |||::: |||:::
97 gGlyAlaThrLeuAlaLeuThrGlnValThrProGlnAspGluArgIleP 114
   |||::: |||::: |||::: |||::: |||::: |||:::
228 CAATTCACGCTGACCTGAGCCCTGTGATGATTTGAGAACGACACATCTT 277
   |||::: |||::: |||::: |||::: |||::: |||:::
114 heLeuCysGln.....GlyLysArgProArgSerGlnGlyTyrArg 127

```



```
Sequence 144, Application US/08184009
Patent No. 5833975
GENERAL INFORMATION:
APPLICANT: Proletti, Enzo
APPLICANT: Tartaglia, James
APPLICANT: Cox, William I.
TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
NUMBER OF SEQUENCES: 217
CORRESPONDENCE ADDRESS:
ADDRESS: Curtis, Morris & Safford
STREET: 530 Fifth Avenue
CITY: New York
STATE: NY
ZIP: 10036
COUNTRY: USA
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/184,009
FILING DATE: 19-JAN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2530
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
TELEX: 425066CURTMS
INFORMATION FOR SEQ ID NO: 144:
SEQUENCE CHARACTERISTICS:
LENGTH: 2434 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-184-009-144

Alignment_scores:
Quality: 242.00 Length: 616
Ratio: 0.910 Gaps: 23
Percent Similarity: 43.182 Percent Identity: 21.104

Alignment_block:
US-09-653-961-2 x US-08-184-009-144 ..

Align seg 1/1 to: US-08-184-009-144 from: 1 to: 2434

91 ArgLeuSerLeuGlnAspArgGlyAlaThrLeuAlaLeuThrGlnValIph 107
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
748 AGCGTGCACATGTCATGCGCAACAGACCGCTCCTCATCTATTCAATGCTAC 797
107 rProGlnAspGluArgGlyLeuPheLeuGlnGlyLeuValArgProArgSer. 123
|:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
798 AAGAAATGACACAGCAAGCTCAATGTGAACCCAGAACCCAGAGAGTG 847
124 ....GlnGlyIrrArgGlnLeuArgValTyrLeuAlaProGluGlu 138
|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
848 CAGGCGCGAGTATCAGTCATCTGAAATGCTCTATGCGCCGGATGCG 897
139 ProAsnIleGlnValAsnProLeuGlyIleProValAsnSerLeuGluPr 155
|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
898 CCCACCATTT.....TCCCCTCTAAACACATCTTACAGATCAGGGGAAA 941
155 oGlnGluValAlaThrCysValGlyArgAsnGlyTyrProIleProGln 172
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
942 TCTGAACCTCTCTGCGACGACGAGCTTAC.....CCACTGCACAGT 985
172 aIleTrrTyrIysAsnGlyArgProLeuLysGlnGluLysAsnArgVal 188
```

```
986 ACTCTGTGTTGTGTCATGG.....ACT 1008
|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
189 HisIleGlnSerSerGln.....ThrValGluSe 198
|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1009 TTCCAGCATCCACCCAGACGCTCTTATCCCAACATCACTAGTGAAATA 1058
198 rSerGlyLeuTyrThrIleGlnSerIleLeuValAlaGlnLeuValLysG 215
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1059 TAGTGATCTTATACGTGCGCAAGCC..... 1083
215 LuAspLysAspAlaGlnPheTyrCysGlnLeuAsnTyrArgLeuProSer 231
|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1084 .....CATAACTCA 1092
232 GlyAsnHisMetLysGluSerArgGluValThrValProValPheTyrPr 248
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1093 GACACTGGCCTCAATAGACCCAGTCACGACGATCAGTCTATGCGAGA 1142
248 oThrGluLysValIrrLeuGluValGluProValGlyMetLeuLysGlu 265
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1143 GCCACCCAAACCCCTTATCACCAGCAACACTCCAAACCCGTGGAGGATG 1192
265 LyAspArgValGluIleArgCysLeuAlaAspGlyAsnProProHis 281
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1193 AGAATGCTGTAGCCTTAACCTGTGACCTGATTCAGAACACAACTAC 1242
282 PheSerIleSerLysGlnAsnProSerThrArgGluAlaGluGluGlu 298
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1243 CTGTGTGGGTAAATATACAGAGCCTCCGCTCAGTCCAGGCTGCAGCT 1292
298 rThrAsnAspAsnGlyValLeuValLeuGluProAlaArgLysGlnHis 315
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1293 GTCCATATACAACAGAGACCCCTACTCTATCAGTGTACCAAGGATATG 1342
315 eArgIrrGlyTyrGluCysGlnGlyLeuAspLeuAspThrMetIleSerLeu 331
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1343 TAGAGACCTTATGAGTGT...GGAATCCAGAACAAATTAAGTGTGAC... 1386
332 LeuSerGluProGlnGluLeuLeuValAsnTyr...ValSerAspValAr 347
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1387 CACAGCGACCCAGTCATCTGAAATGCTCTATGCGCCAGCAGCCACC 1436
347 gValSerProAlaAlaProGluArgGlnGluGlySerSerLeuThrLeu 364
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1437 CATTTCCCTCCATACACATATTACCGTCCAGGCGTAACTCAGCTCT 1486
364 hrCysGluAlaGluSerSerGlnAspLeuGluPheGlnTrrLeuArgGlu 380
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1487 CCTGCCATGACGCTCTAACCCACCTGCACAGATATCTGTGGCTGATTGAT 1536
381 GluThrGlnValLeuGluArgGlyProValLeuGlnIleuHisAspLe 397
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1537 .....GGAGACATCCAGCAACACACAAAGAGCTCTTATTCACACAT 1580
397 uLysArgGluAlaGlnGlyGlyTyrArgCysValAla..... 409
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1581 CACTGAGAGAAGAACAGCGAGCTCTATACCTGCCAGGCCAATTAATCAGCCA 1630
409 ..... 409
1631 GTGGCCACAGCAGACTACAGTCAAGCAATACACAGTCTGTGGGAGCTG 1680
409 ..... 409
1681 CCCAAGCCCTCCATCTCAGACACAACTCCAAACCCGTGGAGACAAGA 1730
409 ..... 409
1731 TGCTGTGGCCTTACACTGTGTAAGCTGAGGCTCAAGAACAACTACTCTGT 1780
410 .....SerValPro.SerIleProGlyLeu..... 417
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
```



```

1084 .....CATAACTCA 1092
232 GYAAsnHISMetLysGluSerArgGluValThrValProValPheTyrPr 248
      ::::::::::::::::::::
1093 GACACTGGCGCTCAATATGAGACACAGACTGACAGACTACAGTCTATGACGA 1142
      ::::::::::::::::::::
248 oThrGluValThrLeuGluValGluProValGluMetLeuLysGlu 265
      ::::::::::::::::::::
1143 GCCACCCAAACCTTCATCATCCAGCAACAACCCCGTGGAGATG 1192
      ::::::::::::::::::::
265 LysAspArgValGluIleArgCysLeuAlaAspGlyAsnProProHis 281
      ::::::::::::::::::::
1193 AGAATGCTGTAGCCTTAACCTGTGAACCTGAGATTGACAAACAACCTAC 1242
      ::::::::::::::::::::
282 PheSerIleSerLysGluAsnProSerThrArgGluAlaGluGlu 298
      ::::::::::::::::::::
1243 CTGTGGTGGTAAATATATCAGAGCTCCGGTCAGTCCAGCGCTGAGCT 1292
      ::::::::::::::::::::
298 rThrAsnAspAsnGlyValLeuValLeuGluProAlaArgLysGluHis 315
      ::::::::::::::::::::
1293 GTCGAATGACACAGAGACCTCACTCACTCACTCACTCACTCACTCACTCA 1342
      ::::::::::::::::::::
315 eArgLysArgTyrGluCysGluGlyLeuAspLysAspThrMetIleSerLeu 331
      ::::::::::::::::::::
1343 TAGGACCCCTATGAGTGT...GGAATCCAGACACAAATTAAGTGTGAC... 1386
      ::::::::::::::::::::
332 LeuSerGluProGluGluLeuLeuValAsnTyr...ValSerAspValAr 347
      ::::::::::::::::::::
1387 CACAGCGACCCAGTCATCTGAATGTCCTCTATGCCCCACAGCACGCCAC 1436
      ::::::::::::::::::::
347 GValSerProAlaAlaProGluValArgGluGlySerSerLeuThrLeu 364
      ::::::::::::::::::::
1437 CATTTCCCTTCATACACTATTACCTCCAGGGGTAACTCAAGCTCTCT 1486
      ::::::::::::::::::::
364 hrcysGluAlaGluSerSerGlnAspLeuGluPheGlnThrPheAspGlu 380
      ::::::::::::::::::::
1487 CCTGGCATGACAGCGCTTAACCCAGCTGACAGTATCTTGCTGATGAT 1536
      ::::::::::::::::::::
381 GluThrGluValLeuGluValArgGlyProValLeuGluInLeuHisAspLe 397
      ::::::::::::::::::::
1537 .....GGGAACATCCAGCAACACACACAGAGCTCTTTATCTCCAAAT 1580
      ::::::::::::::::::::
397 uLysArgGluAlaGluGlyGlyTyrArgCysValAla..... 409
      ::::::::::::::::::::
1581 CACTGAGAAAGACACAGCGAGCTTATACCTGCCAGGCCAAATACACAGCA 1630
      ::::::::::::::::::::
409 ..... 409
1631 GTGGCCACAGCAGACTACAGTCAAGACAAATCAGTCTGTGGAGAGCTG 1680
      ::::::::::::::::::::
409 ..... 409
1681 CCCAAGCCTCATCTCCAGCAACAACCCCGTGGAGACAAAGA 1730
      ::::::::::::::::::::
409 ..... 409
1731 TGCTGTGGCTTCACCTGTGAACCTGAGGCTCAGAACAACCACTACCTGT 1780
      ::::::::::::::::::::
410 .....SerValPro.SerIleProGlyLeu..... 417
      ::::::::::::::::::::
1781 GGTGGTAAATGCTCAGAGCTCTCCAGTACGTCCAGGCTGAGCTGTCC 1830
      ::::::::::::::::::::
417 ..... 417
1831 AATGGCAACAGACCTCACTATTCATGTACAAAGAATAGCAGCAG 1880
      ::::::::::::::::::::
418 .....AsnArgThrGluLeuValAsnValAlaIlePhe 428
      ::::::::::::::::::::
1881 AGCCTATGATGTGGAATCCAGAACTCAGTGAAGTGAACCCGAGTGAAC 1930
      ::::::::::::::::::::
429 GlyProProThrPheAla..... 434
      ::::::::::::::::::::
1931 CAGTCACCTGATGCTCTATGAGCGGACACCCCATCATTTCCCC 1980

```

```

435 .....PheLysGluArgLysValThrValLysGluAsn 446
      ::::::::::::::::::::
1981 CCAGACTCTCTTACCTTTCGGGACCGAAC..... 2010
      ::::::::::::::::::::
446 eValLeuAsnLeuSerCysGluAlaSerGlyHisProAlaProThrIle 462
      ::::::::::::::::::::
2011 .....CTCAACCTTCCTCCACTCGGCTCTCAACCCATCCCGCAGATAT 2055
      ::::::::::::::::::::
463 SerThrAsnValAsnGlyThrAlaSerGluGlnAspGlnAspProGlnAr 479
      ::::::::::::::::::::
2056 TCTTGGCGATCAATGAGATA.....CCGACGA 2084
      ::::::::::::::::::::
479 GValLeuSerThrLeuAsnValLeu...ValThrProGluLeuLeuGlu 495
      ::::::::::::::::::::
2085 ACACACACAAGTCTCTTATCGCCAAATTCAGCCCAATATATACGGA 2134
      ::::::::::::::::::::
495 hArgValGluCysThrAlaSerAsn..AspLeuGlyLysAspThrSer 510
      ::::::::::::::::::::
2135 CC...TATGCTGTGTTGTCTCTACTGCTGCTAGTGGCCGCAATATTC 2181
      ::::::::::::::::::::
511 IleLeuPheLeuGluLeuValAsnLeuThrThrLeuThrProAspSerAs 527
      ::::::::::::::::::::
2182 ATAAGTC..... 2187
      ::::::::::::::::::::
527 nThrThrThrGlyLeuSerThrSerThrAlaSerProHisThrArgAla 544
      ::::::::::::::::::::
2188 .AAGAGCATCAGAGTCTCTCATCTGGAACCTCTCTGCTCTCAGCTG 2236
      ::::::::::::::::::::
544 sAsnSerThrThrGluArgLysLeuProGluProLysArgGlyVal 560
      ::::::::::::::::::::
2237 GGGGCACT..... 2244
      ::::::::::::::::::::
561 ValIleValAlaValIleValCysIleLeuValLeuAlaValLeuVal 577
      ::::::::::::::::::::
2245 .....GTGGCATCATGATGTGAGTGTCTGTT.....GGGCT 2276
      ::::::::::::::::::::
577 aValLeuThrPheLeuTyrLysGlyLysLeuProCysArgArg 592
      ::::::::::::::::::::
2277 TGTCTGTATATGATTTTATCTCAGCA...ATTCTGAGCCCGG 2319
      ::::::::::::::::::::
seq_name: /cgn2.6/prodata/2/lna/sb_comb.seq:us-08-460-736-144
seq_documentation_block:
/ Sequence 144, Application US/08460736
/ Patent No. 5265189
/ GENERAL INFORMATION:
/ APPLICANT: Paoletti, Enzo
/ APPLICANT: Tartaglia, James
/ APPLICANT: Cox, William I.
/ TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
/ NUMBER OF SEQUENCES: 217
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Curtis, Morris & Safford
/ STREET: 530 Fifth Avenue
/ CITY: New York
/ STATE: NY
/ COUNTRY: USA
/ ZIP: 10036
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/460,736
/ FILING DATE: 02-JUN-1995
/ CLASSIFICATION: 514
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/184,009
/ FILING DATE: 19-JAN-1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Frommer, William S.

```



```

479 gValLeuSerThrLeuAsnValLeu...ValThrProGluLeuLeuGluTr 495
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2085 ACACACACAGATCTCTTATGCCCAAAATCAGCCCAAAATATATACGGGA 2134
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
495 hGluValGluCysThrAlaSerAsn...AspLeuGlyLysAsnThrSer 510
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2135 CC...TATGCGCTGTGTTGTCTCTAATCTGCTACTGCGCCGCAATATATCC 2181
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
511 IleLeuPheLeuGluLeuValAsnLeuThrThrLeuThrProAspSerAs 527
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2182 ATAGTC..... 2187
527 nThrThrThrGlyLeuSerThrSerThrAlaSerProHisThrArgAlaA 544
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2188 .AGACATCATCAGCTCTGTCATCTGAGACTCTCTCTGCTCTCAGCTG 2236
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
544 snSerThrSerThrGluArgLysLeuProGluProGluSerArgLysVal 560
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2237 GGGCCACT..... 2244
561 ValIleValAlaValIleValGlySLeuValLeuAlaValLeuGluVal 577
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2245 .....GTGGCATCATGATGAGTGGTGGT.....GGGCT 2276
577 aValLeuTyrrPheLeuTyrrLysGlyLysLeuProCysArgArg 592
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2277 TGCTCTGATATAGTTTATCTTCGAGGA...ATTCCTGCGAGCCCG 2319
seq_name: /cgn2_6/ptodata/2/1na/5B_COMB.seq:US-08-184-009-145

seq_documentation_block:
; Sequence 145..Application US/08184009
; Patent No. 5833975
; GENERAL INFORMATION:
; APPLICANT: Paolletti, Enzo
; APPLICANT: Tartaglia, James
; APPLICANT: Cox, William I.
; TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 217
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/184,009
; FILING DATE: 19-JAN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2530
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEEX: 425066CURTMS
; INFORMATION FOR SEQ ID NO: 145:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2349 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-08-184-009-145

```

alignment_scores:

Quality:	241.50	Length:	611
Ratio:	0.918	Gaps:	23
Percent Similarity:	43.044	Percent Identity:	21.277

alignment_block:

US-09-653-961-2 x US-08-184-009-145 ..

Align seg 1/1 to: US-08-184-009-145 from: 1 to: 2349

```

91 ArgLeuSerLeuGlnAspArgGlyAlaThrLeuAlaLeuThrGlnValTh 107
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
751 AGGCTGCGAGCTGTCCAAATGCGACAGGACCGCTCATCTATTCATATGTCAC 800
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
107 rProGlnAspGluArgIlePheLeuCysGlnGlyLysArgProArgSer. 123
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
801 AAGAAATGACACAGCAGCACTCAAAATGTGAACCCCAACACCACTGAGTG 850
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
124 .....GlnGluTyrrArgIleGlnLeuArgValTyrrLysAlaProGlu 138
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
851 CGAGCGCAGTGAATTCAGTCAATCTGAATGCTCTATGGCCGCGATGCGC 900
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
139 ProAsnIleGlnValAsnProLeuGlyIleProValAsnSerLysGluPr 155
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
901 CCCACCAATT.....TCCCTCTAAGACACATCTTACAGATCGGGGAAA 944
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
155 oGluGluValAlaThrCysValGlyArgAsnGlyTyrrProIleProGlu 172
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
945 TCTGAACTCTTCCTGCGCCAGCGACTTAAC.....CCACCTGACAGT 988
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
172 alIleTyrrLysAsnGlyArgProLeuLysGluGluLysAsnArgVal 188
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
989 ACTCTTGCTTTGTCAATGG.....ACT 1011
189 HisIleGlnSerSerGln.....ThrValGlu 198
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1012 TTCAGCAATCCACCCAGAGCTCTTATCCCAATCCATCGATGATGAATA 1061
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
198 rSerGlyLeuTyrrThrLeuGlnSerIleLeuLysAlaGlnLeuValLysG 215
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1062 TAGTGGATCTATAGCTGCGCAAGCC..... 1086
215 LuAspLysAspAlaGlnPheTyrrCysGluLeuAsnTyrrArgLeuProSer 231
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1087 .....CATACTCA 1095
232 GlyAsnHisMetLysGluSerArgGluValThrValProValPheTyrr 248
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1096 GACACTGGCCCTCAATAGGACACAGCTCAGCAGTCACAGTCTATGACAGA 1145
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
248 OThrGluLysValThrLeuGluValGluProValGlyMetLeuLysGluG 265
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1146 GCCACCAACCAACCTTCATCACCAGCAGCAACCTCAACCCCGTAGAGATG 1195
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
265 LysAspArgValGluIleArgCysLeuAlaAspGlyAsnProProHis 281
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1196 AGGATGCTGTAGCCCTTACCTGTGAACCTGAGATTTCAGAACACAACTAC 1245
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
282 PheSerIleSerLysGlnAsnProSerThrArgGluAlaGluGluThr 298
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1246 CTGTGTGGTGAATAATATCAGAGCCTCCGGTCACTGCCAGCGACGACGT 1295
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
298 rThrAsnAspAsnGlyValLeuValLeuGluProAlaArgLysGluHis 315
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1296 GTCCAATGACAAACAGGACCTCATCTACTCAGTGTCAAGAGATGATG 1345
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
315 eArgLysArgTyrrGluCysGlnGlyLeuAspLeuAspThrMetIleSerLeu 331
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1346 TAGGACCTATGAGTGT...GGAATCCAGAACGATTAAAGTGTGAC... 1399
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
332 LeuSerGluProGlnGluLeuLeuValAsnTyrr...ValSerAspValAr 347
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1390 CACAGCGACCCAGTCATCTGAATGTCTCTATGCGCCAGACGACCCACC 1439

```

```
347 gValserProAla1a1aProGluArgGluGlySerSerLeuThrLeuT 364
      :::::::::::
1440 CATTTCCCTCCCTACACCTATTACCTGTCAGAGCGAACCCTCT 1489
364 hrcysGluAlaGluSerSerGluAspLeuGluPheGluTrpLeuArgLu 380
      :::::::::::
1490 CCTGCCATCAGCCTTACCCACCTGACAGATATCTTGCTGATGAT 1539
381 GluThrGlyGluValLeuGluValArgLysProValLeuGluLeuHisAspLe 397
      :::::::::::
1540 .....GGGACATCCACACACACACAGAGCTTTATCTCCACAT 1583
397 uLysArgGluAlaGlyGlyGlyTyrArgCysValAla..... 409
      :::::::::::
1584 CACTGAGAAAGAACAGCGACTCTATACCTGCCAGGCCAATAACTAGCCCA 1633
409 ..... 409
1634 GTGGCCACAGCAGACTACAGTCACAGACAAATGACAGTCTCGCGAGCTG 1683
409 ..... 409
1684 CCCAAGCCTCCATCTCCAGCAACAACCTCCAAACCGTGAGAGACAGGA 1733
409 ..... 409
1734 TGCTGTGGCCTTCACCTGTGAACCTGAGGCTCAGAACACACACCTTCT 1783
410 .....SerValPro.Ser1LeuProGlyLeu..... 417
      :::::::::::
1784 GGTGGGTAAATGTGTAGAGGCTCCAGTCACTCCAGGCTGCAGCTGTCC 1833
417 ..... 417
1834 AATGGCAACAGAGCCTCACTATTCAATGTCTCAAGAATGACGCAAG 1883
418 .....AsnArgThrGluLeuValAsnValAlaIlePhe 428
      :::::::::::
1884 AGCCTATGTGTGTGAATCCAGAACTCAGTAGTGCACAAACGAGTAGACC 1933
429 GlyProProTrpMetAla..... 434
      :::::::::::
1934 CAGTCACCTCGAGATGCTCTATAGGCGCGGACACCCCATCATTTCCCCC 1983
435 .....PheLysGluArgLysValIlePValLysGluAsnM 446
      :::::::::::
1984 CCAGACTCGTCTTACCTTGGGAGCGAAC..... 2013
446 etValLeuAsnLeuSerCysGluAlaSerGlyHisProArgProThrIle 462
      :::::::::::
2014 .....CTCAACCTCTCTGCCACCTCGGCCCTTACACCATCCCGCAGTAT 2058
463 SerTrpAsnValAsnGlyThrAlaSerGluGlnAspGlnAspProGlnArg 479
      :::::::::::
2059 TCTTGGCGTATCAATGGGATA.....CCGACAGCA 2087
479 gValLeuSerThrLeuAsnValLeu...ValThrProGluLeuLeuLur 495
      :::::::::::
2088 ACACACACAAAGTCTCTTTATCGCCAAATACGCCCAATATATACCGGA 2137
495 hrgLysValGluCysThrAlaSerAsn...AspLeuGlyLysAsnThrSer 510
      :::::::::::
2138 CC...TATGCTGTGTTGTCTCTACACTGGCTACTGGCCGCAATATATCC 2184
511 IleuPheLeuGluLeuValAsnLeuThrThrLeuThrProAspSerAs 527
      :::::::::::
2185 ATAGTC..... 2190
527 nThrThrThrGlyLeuSerThrSerThrAlaSerProHisThrArgAla 544
      :::::::::::
2191 .AAGGATCATCAGTCTCTGATCTGGAACTTCTCTGCTCTCAGCTG 2239
```

```
544 snSerThrSerThrGluArgLysLeuProGluProGluSerArgGlyVal 560
      :::::::::::
2240 GGGCGACT..... 2247
561 ValIleValAlaValIleValCysIleLeuValLeuAlaValLeuGlyAl 577
      :::::::::::
2248 .....GTCCGATCATGATGATGGAGTCTGCTGTT.....GGGGT 2279
577 aValLeu.TyrPheLeuThrLysLysGly 586
      :::::::::::
2280 TGCTGTGATATAGTTTATCTCGAGGGA 2308

seq_name: /cgn2_6/plodata/2/lna/5B.COMB.seq:US-08-458-356-145

seq_documentation_block:
; Sequence 145, Application US/08458356
; Patent No. 5942235
; GENERAL INFORMATION:
; APPLICANT: Paolletti, Enzo
; APPLICANT: Tartaglia, James
; APPLICANT: Cox, William I.
; TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 217
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,356
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/184,009
; FILING DATE: 19-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2530
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 42506CURTMS
; INFORMATION FOR SEQ ID NO: 145:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2349 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-458-356-145

alignment_scores:
      Quality: 241.50      Length: 611
      Ratio: 0.918      Gaps: 23
Percent Similarity: 43.044      Percent Identity: 21.277

alignment_block:
US-09-653-961-2 x US-08-458-356-145 ..

Align seg 1/1 to: US-08-458-356-145 from: 1 to: 2349

91 ArgLeuSerLeuGlnAspArgGlyAlaThrLeuAlaLeuThrGluValTh 107
      :::::::::::
751 AGGCTGACGCTGTCAATGACAAATGACAGACCCCTCATCTATTCAATGTCAC 800
```

[illegible]

397	uLysArgGlnAlaGlyLysIleTyrArgCysValAla.....	409
1584	CAGTGAAGAAGAACGCGAGCTCATTACCTTCCGAGCCAAATAATCATGACCA	1633
409	409
1634	GTGGCCACACAGAGAGACTACAGTCAACAGCAATCATCAAGTCTCTCGGAGCTG	1683
409	409
1684	CCCAAGCCCTTCATCTTCACAGCAACTCCAAACCGGTGGAGACAAGGA	1733
409	409
1734	TGCTGTGGCCTTACCTGTGTAACCTAGAGCTAGAACACACTTACTGT	1783
410SerValPro.SerIleProGlyLeu.....	417
1784	GGTGGTAAATGGTTCAGAGGCTCCAGTCAAGTCCAGAGGTGACGTGCC	1833
417	417
1834	AATGGCAACGAGACCTCAGTCTATTCAATGTACAGAAATGACGCAAG	1883
418AsnArgThrGlnLeuValAlaValAlaIlePhe	428
1884	AGCGTATGTATGTGGAAATCCAGAACTCAAGTCAAGTGCMAACCGCAGTACC	1933
429	GlyProProTrpMetAla.....	434
1934	CAGTCAACCTGTGATGTCTCATATGGGCGCGACACCCCATATTTCGCC	1983
435PheLysGlnArgLysValTrpAllyLysGlnM	446
1984	CCAGACTCGCTTACTCTTCGGGAGCGAAC.....	2013
446	eValLeuAsnLeuSerCysGlnAlaSerGlnHisProArgProThrIle	462
2014CTCAACCTCTCTGCGACCTGCGCCTCAACCCATCCCGCAGAT	2058
463	SerTrpAsnValAsnGlyThrAlaSerGlnAspGlnAspProGlnArg	479
2059	TCTTGGCGATACATGGGATA.....CCGACGA	2087
479	GValLeuSerThrLeuAsnValLeu...ValThrProGluLeuLeuGlu	495
2088	ACACACACAAGTCTCTTATATGCCMAAATCCGCAAAATAAATAACGGGA	2137
495	hrcLysAlaGlySerThrAlaSerAsn...AspLeuGlyLysAsnThrSer	510
2138	CC...TATCCCTGTTTGTCTCTTAACCTTGAGTACGGCGCAATAAATATCC	2184
511	IleLeuPheLeuGluIleuValAsnLeuThrThrLeuThrProAspSerAs	527
2185	ATAGTC.....	2190
527	nhrThrThrGlyLeuSerThrSerThrAlaSerProHisThrArgAlaA	544
2191	AAGACACATCAAGTCTGTGATCTGGAATCTTCCTGGTCTCTCAAGCTG	2239
544	snSerThrSerThrGlnArgLysIleuProGluProGluSerArgGlyVal	560
2240	GGGCGACT.....	2247
561	ValIleValAlaValIleValCysIleLeuValLeuAlaValLeuGlnLysI	577
2248GTGGCATCATGATTTGAGTGGTGGTT.....GGGCT	2279
577	aValLeu.TyrPheLeuTyrIlylSylSgly	586
2280	TGCTCTGATATAGTTTATCTTCAGAGGA	2308

seq_name: /cgn2_6/ptodata/2/lna/68_COMB.seq:US-08-460-736-145


```

1734 TCGTGGCCCTTCACCTGTGTAACCTGAGGCTCAGAACACAACTACCTGT 1783
410 .....SerValPro.SerIleProGluLeu..... 417
1784 GGTGGGTAATGTGTGACAGGCTCCAGTGTCCAGGCTGCAGCTGTCC 1833
417 ..... 417
1834 AATGGCAACAGACCTCACTATTCATATGTCACAAAGAAATGAGCCAG 1883
418 .....AsnArgThrGlnLeuValAsnValAlaIlePhe 428
1884 AGCCTGTATGTGTAATCAGAACTCACTAGTGTGCAAAACCCAGTGTACC 1933
429 GlyProProThrPheAla..... 434
1934 CAGTACACCTGGATGTCTATGGCCGCGACACCCCATTCATTCGCC 1983
435 .....PheLysGluArgLysValTyrValLysGluAsnM 446
1984 CCAGACTCGCTTACCTTTCGGGAGCGAAC..... 2013
446 eValLeuAsnLeuSerCysGluAlaSerGlyHisProArgProThrIle 462
2014 .....CTCAACTCTCTCTCCCTCGGCTTACCCCATCCCGCAGTAT 2058
463 SerTyrPasnValAsnGlyThrAlaSerGluAsnGlnAspProGlnArg 479
2059 TCTTGGCGGTATCAATGGGATA.....CCGCACGA 2087
479 gValLeuSerThrIleAsnValLeu...ValThrProGluLeuLeuGluT 495
2088 AACACACACAGATTCTTATTCGCCAAATATCACGCCAAATATATACGGGA 2137
495 hGlyValGluCysThrAlaSerAsn...AspLeuGlyLysAsnThrSer 510
2138 CC...TATGCCGTGTTGTCTCTACTGTCGCGCTGAGCCGCGCAATATATCC 2184
511 IleLeuPheLeuGluLeuValAsnLeuThrThrProAspSerAs 527
2185 ATAGTC..... 2190
527 nThrThrThrGlyLeuSerThrSerThrAlaSerProHisThrArgAla 544
2191 AAGAGCATCACAGTCTCTGCTGAGACTTCTCTGCTGCTCAGAGTCC 2239
544 snSerThrSerThrGluArgLysLeuProGluProGluSerArgGlyVal 560
2240 GGGCCACT..... 2247
561 ValIleValAlaValIleValCysIleLeuValLeuAlaValLeuGlyAl 577
2248 .....GTGCGCATCATGATTGTGAGTGTCTGTT.....GGGGT 2279
577 aValLeu.TyrPheLeuTyrLysLysGly 586
2280 TGCCTGTATATAGTTTATCTCAGAGGA 2308
seq.name: /cgn2.6/ptodata/2/lna/5b_comb.seq:us-08-602-725-35
seq_documentation_block:
; Sequence 35, Application US/08602725
; Patent No. 5965710
; GENERAL INFORMATION:
; APPLICANT: BODMER, WALTER F
; APPLICANT: DUBBIN, HELGA
; APPLICANT: SNARY, DAVID
; APPLICANT: STEWART, LORNA MD
; APPLICANT: YOUNG, SUSAN
; APPLICANT: BATES, PAUL A
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES FOR USE IN
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF COLORECTAL CANCER

```

```

; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/602,725
; FILING DATE: 02-FEB-1996
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB94/01816
; FILING DATE: 19-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9317423
; FILING DATE: 21-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B.J.
; REGISTRATION NUMBER: 36663
; REFERENCE/DOCKET NUMBER: 1090-8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-816-4091
; TELEFAX: 703-816-4100
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2097 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Human carcinoembryonic antigen
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2094
; US-08-602-725-35

alignment_scores:
Quality: 241.00 Length: 557
Ratio: 0.984 Gaps: 21
Percent Similarity: 43.986 Percent Identity: 22.083

alignment_block:
US-09-653-961-2 x US-08-602-725-35 ..
Align seg 1/1 to: US-08-602-725-35 from: 1 to: 2097

91 ArgLeuSerLeuGlnAspArgGlyAlaThrIleAlaLeuThrGlnValTh 107
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
556 AGCGTCAGCTGTCGAATGCAAGCAAGACCGCTCAGCTTCAATGTCC 605
107 rProGlnAspGluArgIlePheLeuGlnGlyLysArgProArgSer. 123
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
606 AAGAAATGACACAGCAAGCTCAAAATGTCAACCCAGAACCCAGTAGTG 655
124 .....GlnGluTyrArgIleGlnLeuArgValTyrLysAlaProGluG 138
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
656 CCAGGCGCAGTATTCATCTGATGATGCTCTATGGCCCGGATGCC 705
139 ProAsnIleGlnValAsnProLeuGlnIleProValAsnSerLysGluPr 155
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
706 CCACCATTT.....TCCCTCTAAACACATCTTACAGATCAGGGGAAA 749
155 oGluGluValAlaThrCysValGlyArgAsnGlyTyrProIleProGlnV 172

```



```

1539 TGCATGGCCCTTCACTGTAAGCACTGAGGCTCAGAACACAACTACTGTT 1588
410 .....SerValPro.SerLieProGlyLeu..... 417
1589 GGTGGGTAAATGGTCAGAGCCCTCCAGTAGTCCAGGCTGAGCTGTCC 1638
417 ..... 417
1639 AATGCACAGACCACTCACTATTCAATGTGCACAAAGAAATGACCAAG 1688
418 .....AsnArgThrGlnLeuValAsnValAlaIlePhe 428
1689 AGCCTATGATGTGAATCAGAACTCAGATGATGCAAAACCGAGTGACC 1738
429 GlyProProThrMetAla..... 434
1739 CAGTCACTCGATGATTCCTCTATGGCCCGGACACCCCATCAATTTCCCC 1788
435 .....PheLysGlnArgIysValITrpValLysGlnAsn 446
1789 CCAGACTGCTTACTCTTCGGAGGAGCA..... 1818
446 etValLeuAsnLeuSerGlyAlaIleSerGlyHisProArgProThrIle 462
1819 .....CTCAACCTCTCTCGCACTGCGACCTCAACCATCCCGCAGTAT 1863
463 SerTrpAsnValAsnGlyThrIleAlaSerGlnGlnAspGlnAspProGlnArg 479
1864 TCTTGCGCATCATATGGATA.....CCGACAGA 1892
479 gValLeuSerThrLeuAsnValLeu...ValThrProGlnLeuLeuGln 495
1893 ACACACACAAAGTTCTCTTATATCCCAAAATCACGCCAAATATATACGGGA 1942
495 hngIyValGlyCysThrIleAsnSerAsn...AspLeuGlyLysAsnThrSer 510
1943 CC...TATGCTGTGTTTGTCTTAACCTGGCTACTGGCCGCAATATTC 1989
511 IleLeuPheLeuGlnLeuValAsnLeuThrThrLeuThrProAspSerAs 527
1990 ATTA.....GTCAAGACGATCACAGTCTGTCA...TCTGG 2021
527 nThrThrThrGlyLeuSer 533
2022 AACTCTCTGTCGTCTCA 2040

seq_name: /cgn2_6/plodata/2/ina/5A_COMB.seq:US-08-389-459A-16
seq_documentation_block:
; Sequence 16, Application US/08389459A
; Patent No. 3617512
; GENERAL INFORMATION:
; APPLICANT: Morrow, Casey D. and Porter, Donna, C.
; TITLE OF INVENTION: ENCAPSIDATED RECOMBINANT POLIOVIRUS
; TITLE OF INVENTION: NUCLEIC ACID AND METHODS OF MAKING AND
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, SUITE 510
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/389,459A
; FILING DATE: 15-FEB-1995
; CLASSIFICATION: 435

```



```
446 etValLeuAsnLeuSerCysGluAlaIleSerGlyHisProArgProThrIle 462
      |||||||
1927 .....CTCAACCTCTCTCGCTGACCTCGGCTCTAACCACGCCGCGAGTAT 1971
      |||||||
463 SerTrpAsnValAsnGlyThrAlaIleSerGluGlnAspGlnAspProGlnAr 479
      |||||||
1972 TCTGGCGGTATCAATGGATA.....CCGCAGCA 2000
      |||||||
479 gValLeuSerThrLeuAsnValLeu...ValThrProGluLeuLeuGlnT 495
      |||||||
2001 ACACACACAGATCTCTTATTCGCCAAATACGCCCAATATATACGGGA 2050
      |||||||
495 hrgIyValGlyCysThrAlaSerAsn...AspLeuGlyLysAsnThrSer 510
      |||||||
2051 CC...TATGCCCTGTTGTCTTAAGTGGCTACTGGCGGCAATATATTC 2097
      |||||||
511 IleLeuPheLeuLeuValAsnLeuThrThrLeuThrProAspSerAs 527
      |||||||
2098 ATA.....GTCAAGACATACAGTCTCTGCA...TCTGG 2129
      |||||||
527 nThrThrThrGlyLeuSer 533
      |||||||
2130 AACTCTCCGCTCTCTCA 2148

seq_name: /cgn2.6/ptodata/2/ina/6A_COMB.seq:US-08-987-867A-16

seq_documentation_block:
: Sequence 16, Application US/08987867A
: Patent No. 6063384
: GENERAL INFORMATION:
: APPLICANT: C. Morrow et al.
: TITLE OF INVENTION: ENCAPSULATED RECOMBINANT VIRAL
: TITLE OF INVENTION: NUCLEIC ACID AND METHODS OF MAKING AND
: NUMBER OF SEQUENCES: 23
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: LAHIVE & COCKFIELD
: STREET: 28 STATE STREET
: CITY: BOSTON
: STATE: MASSACHUSETTS
: COUNTRY: USA
: ZIP: 02109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: ASCII
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/987,867A
: FILING DATE: 09-DEC-1997
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/087,009
: FILING DATE: 01-JUL-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Myers, Louis
: REGISTRATION NUMBER: 35,965
: REFERENCE/DOCKET NUMBER: DAP-004CPDV
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 227-7400
: TELEFAX: (617) 742-4214
: INFORMATION FOR SEQ ID NO: 16:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2220 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..2203
: US-08-987-867A-16
```

```
alignment_scores:
      Quality: 241.00      Length: 557
      Ratio: 0.984      Gaps: 21
      Percent Similarity: 43.986      Percent Identity: 22.083

alignment_block:
US-09-653-961-2 x US-08-987-867A-16 ..
Align seg 1/1 to: US-08-987-867A-16 from: 1 to: 2220

91 ArgLeuSerLeuGlnAspArgGlyAlaThrLeuLeuAlaLeuThrGlnValTh 107
|||||
664 AGCGTCGACGCTCTCAATGCAACAGACGCTCTCACTATTCATTAATGTCAC 713
|||||
107 rProGlnAspGluArgGlyLeuLeuGlyLysArgProArgSer 123
|||||
714 AAGAAATGACACAGCAAGCTACAAATGGAACCCAGAACCCAGTGAAGTG 763
|||||
124 ....GlnGluTyrArgGlyLeuLeuArgValTyrLysAlaProGluGln 138
|||||
764 CCAGCGCAGATGATTCAGTCAATCCGAATGTCCTATGCGCCGAGATGCC 813
|||||
139 ProAsnIleGlnValAsnProLeuGlyTyrProValAsnSerLysGluPr 155
|||||
814 CCACACAT.....TCCCTCTAAACACATCTTACAGATCAGGGAATA 857
|||||
155 oGluGluValAlaThrCysValGlyArgAsnGlyTyrProIleProGlu 172
|||||
858 TCTGAACCTCTCTGCCATGCAGCCTCTAAC.....CCACCTGCACAGT 901
|||||
172 alleIrrPyrLysAsnGlyArgProLeuLysGluGluLysAsnArgVal 188
|||||
902 ACTCTGCTTGTCTCAATGG.....ACT 924
|||||
189 HisIleGlnSerSerGln.....ThrValGlnSer 198
|||||
925 TTCACGCAATCCACCAAGACCTCTTATCCCAACATCACTGAGATTA 974
|||||
198 rSerGlyLeuTyrThrLeuGlnSerIleLeuLysAlaGlnLeuValLysG 215
|||||
975 TAGTGATCTTATACGTGCCAAGCC..... 999
|||||
215 LuAspLysAspAlaGlnPheTyrCysGluLeuAsnTyrArgLeuProSer 231
|||||
1000 .....CATACTCA 1008
|||||
232 GlnAsnHisMetLysGluSerArgGluValThrValProValPheTyrPr 248
|||||
1009 GACACTGGCTCTCAATAGACACAGTCACGACGATACAGTCATATGACGA 1058
|||||
248 cThrGlnLysValThrPheLeuGluValGluProValGlyMetLeuLysGlu 265
|||||
1059 GCCACCCCAACCTTCATCACACGACCAACCTCCACCCGCTGGAGGATG 1108
|||||
265 LysAspArgValGluIleArgCysLeuAlaAspGlyAsnProProGlnHis 281
|||||
1109 AGAATGCTGATGACCTTAACCTGTGAACCTGATTCGAAACACACATAC 1158
|||||
282 PheSerIleSerLysGlnAsnProSerThrArgGluAlaGluGluGluTh 298
|||||
1159 CTGTGGTGGGTAAATATCAAGAGCTCCCGGTACGTCCAGGCTGACGT 1208
|||||
298 rThrAsnAspAsnGlyValLeuValLeuGluProAlaArgLysGlnHis 315
|||||
1209 GTCCAAATGACACAGACGCTCACTCACTCACTGATGCAAGAAATGATG 1258
|||||
315 erGlyArgTyrGlyCysGlnGlyLeuAspLeuAspThrMetIleSerLeu 331
|||||
1259 TAGGACCTTATGAGTGT...GGAATCCAGAACGATTAAGTGTGAC... 1302
|||||
332 LeuSerGluProGlnGluLeuLeuValAsnTyr...ValSerAspValAr 347
|||||
```

```
1303 CACAGGACCCAGTATCTATGTCCTTATGGCCGAGACGCCAC 1352
347 gValSerProAlaIaIaProGluIaGluIaGluIaSerLeuThrLeuT 364
      :::::::::::::: :::: :::: :::: :::: ::::
1353 CATTTCCCTCCCTATACACCTATTATCCAGGGGAGAACCTCACCTCT 1402
364 hCysGluIaGluSerSerGluAspLeuPheGlnIaPheGlnIaGlu 380
      :::::::::::::: :::: :::: :::: :::: ::::
1403 CCGCATGACAGCTTAAACCCAGCCGACATATTTCTGGGTGATGAT 1452
381 GluThrGluIaValLeuGluIaGluIaProValLeuGluIaAspLe 397
      :::::::::::::: :::: :::: :::: :::: ::::
1453 .....GGGAAAGATCCAGCAACACACAAAGAGCTTTATCTCCAAAT 1496
397 uLysArgGluIaGluIaGluIaGluIaGluIaGluIaGluIaGlu 409
      :::::::::::::: :::: :::: :::: :::: ::::
1497 CACTGAGAAAGAACAGGAGCTCTATACCTGCGAGGCAATAACAGCCA 1546
409 ..... 409
1547 GTGGCCACAGAGAGACTACAGTACAGACATACAGTCTCTCGGAGCTG 1596
409 ..... 409
1597 CCCAAGCCCTCATCTCCAGCAACACTCCAAACCCGTGAGAGACAGGA 1646
409 ..... 409
1647 TGCTGTGGCCTTCACCTGTGAACCTGAGCCTCAGAAACACACCTTACT 1696
410 .....SerValPro.SerIleProGluLeu..... 417
1697 GGTGGGTAAATGTCAGAGGCTCCAGTCCAGGCTGCAAGCTGTCC 1746
417 ..... 417
1747 AATGGCAACAGAGACCTCACTATTCAATGTCCACAGAAATGACGCAAG 1796
418 .....AsnArgThrGluLeuValAsnValAlaIaIaLeuPhe 428
1797 AGCCTATGTATGTGAATCCAGAACTCAGTGAAGTGCAGAACCCGAGTACC 1846
429 GlyProProThrPheAla..... 434
1847 CACTCAACCCGAGATGCTCTGTATGGCCGAGACACCCCATATTTCGCC 1896
435 .....PheLysGluArgLysValITrPValLysGluAsnM 446
1897 CCAGACTCGTTTACCTTCGAGGAGCAAC..... 1926
446 etValLeuAsnLeuSerCysGluAlaSerGlyHisProArgProThrIle 462
1927 .....CTCAACCTCTCTGCGACACTGCGCCCTTAACCCATCCCGCAGTAT 1971
463 SerTrpAsnValAsnGlyThrAlaSerGluIaAspGlnAspProGlnAr 479
1972 TCTTGGCGTATCATGAGATA.....CCGACGCA 2000
479 gValLeuSerThrLeuAsnValLeu...ValThrProGluLeuLeuGluT 495
2001 ACACACACAGATTCTCTTATCGCCCAAAATACGCCAAATATATACGGCA 2050
495 hrcGlyValGluCysThrAlaSerAsn...AspLeuGlyLysAsnThrSer 510
2051 CC...TATGCTGTGTTTGTCTCTAATCTGTGCTACTGCGCCCATATATTC 2097
511 IleLeuPheLeuGluLeuValAsnLeuThrThrLeuThrProAspSerAs 527
2098 ATA.....GTCAAGAGCATCACAGTCTCTCA...TCTGG 2129
527 nThrThrThrGlyLeuSer 533
2130 AACTTCTCTGTGTTCTTCA 2148
```

```
seq_name: /cgn2_6/prodata/2/lna/6A_COMB.seq:us-08-468-856B-5
seq_documentation_block:
; Sequence 5, Application US/08468856B
; Patent No. 6013772
; GENERAL INFORMATION:
; APPLICANT: Barnett, Thomas; Eiting, James; Kamarek, Michael;
; APPLICANT: Kretschmer, Axel
; TITLE OF INVENTION: CDNAS CODING FOR MEMBERS OF THE
; TITLE OF INVENTION: CARCINOEMBRYONIC ANTIGEN FAMILY
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Sprung Horn Kramer & Woods
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 2.0 Mb storage
; COMPUTER: APPLE MACINTOSH 6500
; OPERATING SYSTEM: SYSTEM 7.5
; SOFTWARE: WordPerfect 3.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,856B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/027,974
; FILING DATE: 08-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/760,031
; FILING DATE: 13-SEP-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/274,107
; FILING DATE: 21-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/207,678
; FILING DATE: 16-JUN-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/060,031
; FILING DATE: 19-JUN-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/016,683
; FILING DATE: 19-FEB-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/896,361
; FILING DATE: 13-AUG-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurt G. Briscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: MDI 242.10-KGB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2839 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-468-856B-5
```

```
alignment_scores:
    Quality: 238.00      Length: 557
    Ratio: 0.971         Gaps: 21
Percent Similarity: 43.986      Percent Identity: 21.903
```

```
alignment_block:
US-09-653-961-2 x US-08-468-856B-5 ..
```

Align seq 1/1 to: US-08-468-856B-5 from: 1 to: 2839

```

91 ArgLeuSerLeuGlnAspArgGlyAlaThrLeuAlaLeuThrGlnValThr 107
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
572 AGCTTCAGCTCCATCCATAGGACACAGACCCCTCATCTATTCATAGTAC 621
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
107 rProGlnAspGluArgIlePheLeuGlyGlnGlyLysArgProAlaSer 123
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
622 AAGAATGAACAAGCAAGCTCAATATGGAACCCAGAACCCAGTAGAG 671
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
124 .....GlnGluThrArgGlnIleGlnLeuArgValThrLysAlaProGlu 138
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
672 CAGGGGCAAGTATTCATCATCTGATGATGCTCTATGAGCCCGAGATGC 721
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
139 ProAsnIleGlnValAsnProLeuGlyIleProValAsnSerLysGlu 155
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
722 GCGACCAAT.....TCCCTCTAACAACATCTTACAGATCAGGGGAA 765
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
155 OGluGluValAlaThrCysValGlyArgAsnGlyTyrProIleProGlu 172
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
766 TGTGAACCTCTCTCCAGCAGCAGCTCTAAC.....CCACCTGCACAG 809
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
172 AlIleTyrTyrLysAsnGlyArgProLeuLysGluLysAsnArgVal 188
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
810 ACTCTGTGTGTCAATGGG.....ACT 832
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
189 HisIleGlnSerGln.....ThrValGlu 198
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
833 TTCCACCAATCCACCCAGAGCTCTTATCCCAACATCAGTGTGATATA 882
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
198 rSerGlyLeuTyrThrLeuGlnSerIleLeuLysAlaGlnLeuValLys 215
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
883 TAAGTGAATCTATACGTGCAGAACCC..... 907
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
215 LysAspLysAlaGlnPheTyrCysGluLeuAsnTyrArgLeuProSer 231
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
908 .....CATACCTCA 916
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
232 GlyAsnHisMetLysGluSerArgGluValThrValProValPheTyr 248
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
917 GACAGTGGCTCATATAGACACAGTCCAGTCCAGTCACTATATCCAGA 966
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
248 orthGlyLysValTyrLeuGluValGluProValGlyMetLeuLysGlu 265
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
967 GCCACCCAAACCTTCATACACAGACACACCTCCAAACCCGTGGAGGATG 1016
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
265 LysAspArgValGluIleArgCysLeuAlaAspGlyAsnProProPhe 281
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1017 AGGATGCTGTAGCTTAACTGTGACCTGTGAGATTCAGAACCAACACTAC 1066
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
282 PheSerIleSerLysGlnAsnProSerThrArgGluAlaGlnGluGlu 298
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1067 CTGTGGTGGGTAAATATATATACAGCCCTCCGCTCAGTCCAGGCTGAGCT 1116
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
298 rThrAsnAspAsnGlyValLeuValLeuGluProAlaArgLysGlnHis 315
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1117 GTCGAATGACAAAGAGACCTCTACTACTACTGATGTCACAGGAATGATG 1166
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
315 erGlyArgTyrLysGlnGlnLysLeuAspLysPheMetIleSerLeu 331
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1167 TAGAGCTATATAGTGT.....GGAATCCAGAACGATTAAGTGTGAC... 1210
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
332 LeuSerGluProGlnGluLeuLeuValAsnTyr...ValSerAspVal 347
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1211 CAAAGGAGACCCAGTCACTCATATGCTCTATGAGCCAGACAGACCCAC 1260
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
347 gValSerProAlaAlaProGluArgGlnGluLysSerSerLeuThrLeu 364
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1261 CATTTCCCTCATACACTATTTACCGTCCAGGGGTCAACCTCAGCTCT 1310
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
364 hTyrGluAlaGluSerSerGlnAspLeuGluPheGlnTyrLeuValGlu 380
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1311 CTGGCATTGACGCTCTTAACCACTGCACAGTATCTTGCTGATGTAT 1360

```

```

381 GluThrGlyGlnValLeuGluArgGlyProValLeuGlnLeuHisAsp 397
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1361 .....GGGAAATCCACAGCAACACACAGAGAGCTCTTATCTCCACAT 1404
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
397 uLysArgGluAlaGlyGlyGlyTyrArgCysValAla..... 409
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1405 CACTGAGAAGAACAGGAGCTCTATACCTGCGAGCAATATACAGCA 1454
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
409 ..... 409
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1455 GTGGCCACAGCAGGACTCAGTCAAGTCAATCAATCACTCTCTGCGAGCTG 1504
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
409 ..... 409
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1505 CCCAAGCCCTCATCTCCAGCAACACTCCAAACCCGTGGAGACAAAGA 1554
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
409 ..... 409
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1555 TGCTGTGGCTTCACTGTGAACTGAGGCTCAGAAACAACTACCTGT 1604
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
410 .....SerValPro.SerIleProGlyLeu..... 417
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1605 GGTGGTAATGTGTAGAGCTCCAGTCAAGTCCAGGCTGAGCTGTCC 1654
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
417 ..... 417
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1655 AATGCAACAGAGCCCTCATCTATTCATATGTCACAAATGACCGCAAG 1704
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
418 .....AsnArgThrGluLeuValAsnValAlaIlePhe 428
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1705 AGCTTATGTATGTGAATCCAGAACTAGTAGTGAAGTCAAAACCGCAGTGC 1754
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
429 GlyProProTyrMetAla..... 434
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1755 CAGTCACCTGGATGTCTCTATGAGCGGACACCCCACTATTCCCC 1804
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
435 .....PheLysGluArgLysValTyrPValLysGluAsn 446
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1805 CCAAGCTCTTACCTTTCGGAGCGAAC..... 1834
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
446 eValLeuAsnLeuSerCysGluAlaSerGlyHisProArgProThrIle 462
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1835 .....CTCAACCTCTCCGACACTGCGCTCTAACCACATCCCGGAGTAT 1879
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
463 SerTrpAsnValAsnGlyThrAlaSerGluGlnAspGlnAspGlnArg 479
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1880 TCTTGGCTATCAATGGATA.....CCGACAGA 1908
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
479 gValLeuSerThrLeuAsnValLeu...ValThrProGluLeuGlu 495
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1909 ACACACACAAGTCTCTTATCGCCAAATCAACCCAAATATATACGGA 1958
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
495 hGlyValGlnCysThrAlaSerAsn...AspLeuGlyLysAspThrSer 510
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1959 CC...TATGCTGTGTCTCTACTTGGCTACTGCGCCGCAATATATCC 2005
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
511 IleLeuPheLeuGlnLeuValAsnLeuThrThrLeuThrProAspSer 527
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2006 ATA.....GTCAAGAGCATCAAGTCTGTGCA...TCGG 2037
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
527 nThrThrThrGlyLeuSer 533
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2038 AACTCTCTCTGCTCTCA 2056
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

seq_name: /cgn2_6/ptodata/2/ina/fa_COMB.seq:US-08-468-859A-5

seq_documentation_block:

Sequence 5, Application US/0846859A

Patent No. 6022958

GENERAL INFORMATION:

APPLICANT: Barnett, Thomas; Eiting, James; Komarck, Michael;

APPLICANT: Kretschmer, Axel

```

? TITLE OF INVENTION: CDNAS CODING FOR MEMBERS OF THE
? NUMBER OF SEQUENCES: 14
? CORRESPONDENCE ADDRESSES:
? ADDRESSER: Sprung Horn Kramer & Woods
? STREET: 660 White Plains Road
? CITY: Tarrytown
? STATE: New York
? COUNTRY: USA
? ZIP: 10591-5144
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
? COMPUTER: APPLE MACINTOSH 6500
? OPERATING SYSTEM: SYSTEM 7.5
? SOFTWARE: Wordperfect 3.5
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/468,859A
? FILING DATE: 06-JUN-1995
? CLASSIFICATION: 435
? PRIOR APPLICATION NUMBER:
? APPLICATION NUMBER: US 08/027,974
? FILING DATE: 08-MAR-1993
? PRIOR APPLICATION NUMBER: US 07/760,031
? FILING DATE: 13-SEP-1991
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/274,107
? FILING DATE: 21-NOV-1988
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/207,678
? FILING DATE: 16-JUN-1988
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/060,031
? FILING DATE: 19-JUN-1987
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/016,683
? FILING DATE: 19-FEB-1987
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 06/896,361
? FILING DATE: 13-AUG-1986
? ATTORNEY/AGENT INFORMATION:
? NAME: Kurt G. Briscoe
? REGISTRATION NUMBER: 33,141
? REFERENCE/DOCKET NUMBER: MDI 242.9-KGB
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (914) 332-1700
? TELEFAX: (914) 332-1844
? INFORMATION FOR SEQ ID NO: 5:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2839 nucleotides
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
?
? US-08-468-859A-5
?
? alignment_scores:
? Quality: 238.00 Length: 557
? Ratio: 0.971 Gaps: 21
? Percent Similarity: 43.986 Percent Identity: 21.903
?
? alignment_block:
? US-09-653-961-2 x US-08-468-859A-5 ..
?
? Align seq 1/1 to: US-08-468-859A-5 from: 1 to: 2839
?
? 91 ArgLeuSerLeuGlnAspArgGlyAlaThrLeuAlaLeuThrGlnValTh 107
? |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
? 572 AGGCTCGACGCTGCTCCATGCGACAGACGACCTCATCTATTCATGTCAC 621
? 107 rProGlnAspArgGlyIlePheLeuGlyGlnGlyLysArgProGlySer 123
? |:|||||:|||||:|||||:|||||:|||||:|||||:|||||
? 622 AGAATATGACACAGCAGCTACAAATGTGAACCCAGAACCCAGTGAGTG 671

```

```

124 ....GlnGluTyrArgIleGlnLeuArgValTyrLysAlaProGlnGlu 138
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
672 CCAAGGCGCAGATGATTCAGTCAATGCTGAAATGCTCTGAGCCCGGAGTCC 721
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
139 ProAsnIleGlnValAsnProLeuGlyIleProValAsnSerLysGluPro 155
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
722 CCCACCAATT.....TCCCTCTAAACACATCTTACAGATGAGCGGAAAT 765
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
155 OGluGluValAlaThrLysValGlyArgGlnGlyTyrProLeuProGln 172
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
766 TCGAACCTCTCTCCGACGAGCGCTCTAAC.....CCAGCTGCACAGT 809
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
172 alIleThrTyrLysAsnGlyArgProLeuLysGlnGluLysAsnArgVal 188
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
810 ACCTGTGGTTGTGCATGG.....ACT 832
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
189 HisIleGlnSerSerGln.....ThrValGlnSer 198
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
833 TTCACCAATCCACCCACAGAGCTTTATCCCAACATCACTGTGAATTA 882
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
198 rSerGlyLeuTyrThrLeuGlnSerIleLeuLysAlaGlnLeuValLys 215
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
883 TAGTGATGCTATGATGTCGCAAGCC..... 907
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
215 LysAspLysAspAlaGlnPheTyrCysGluLeuAsnTyrArgLeuProSer 231
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
908 .....CATAACTCA 916
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
232 GlyAsnHisMetLysGlySerArgGlyValAlaThrValProValPheTyr 248
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
917 GACACTGGGCTCAATAGGACACAGCTCAGACGATCAGACGCTGTGACAA 966
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
248 oThrGlyLysValThrLeuGlnValGlyProValGlyMetLeuLysGlu 265
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
967 GCGACCCAAAGCTTCATGACACAGCAACATCCACCCGAGGAGAGT 1016
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
265 LysAspArgValGlnIleArgCysLeuAlaAspGlyAsnProProGlnHis 281
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
1017 AGATATGCTGTACCTTAACCTGTGAACCTGAGATTCACACACAACTGAC 1066
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
282 PheSerIleSerLysGlnAsnProSerThrArgGlnAlaGlnGluThr 298
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
1067 CTGTGTGGTGAATTAATTAATCAGAGCTCCCGGTCAGCTCCAGGCTGACG 1116
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
298 rThrAsnAspAsnGlyValLeuValLeuGlnProAlaArgLysGlnHis 315
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
1117 GTCCAAATGACAAACAGACGCTCACTACTGCTGCTGACAAAGAAATGAG 1166
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
315 eArgLysArgTyrGlnCysGlnGlyLeuAspLeuAspThrMetIleSerLeu 331
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
1167 TAGGACCTATGAGTGT...GGAAATCAGAACGAAATTAAGTGTGAC... 1210
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
332 LeuSerGluProGlnGluLeuLeuValAsnTyr...ValSerAspValAr 347
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
1211 CACAGCGACCCAGTCACTGTAATGCTGTATGAGCCCAACAGCCAC 1260
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
347 ValSerProAlaAlaProGlnArgGlnGlnLysSerLeuThrLeuThr 364
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
1261 CATTTCCTCCCTCATACACCTTTATTCCTGACGGGTGAACTCAGCTCT 1310
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
364 hCysGluAlaGlnSerSerGlnAspLeuGlnGluProGlnThrLeuArgGlu 380
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
1311 CCTGCCATGACAGCTCTAACCCAGCTGACAGATTTCTGCTGATTTGAT 1360
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
381 GluThrGlnGlnValLeuGlnArgGlyProValLeuGlnLeuHisAspLe 397
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
1361 .....GGGAACATCCACGACACACACACAGAGCTTTTATTCACCACT 1404
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
397 LysArgGlnAlaGlnGlyLysArgCysValAla..... 409
? |||||:|||||:|||||:|||||:|||||:|||||:|||||
1405 CACTGAGAGACAGCGGACTTATACCTGCCAGGCGCAATTACTACGCA 1454

```

```

409 ..... 409
1455 GTGGCCACAGCAGACTACAGTCAAGACATCAAGTCTGTGGGACGTG 1504
409 ..... 409
1505 CCCAAGCCCTCCATCTCCAGCAACACCCAAACCCCTGAGSAGCAAGA 1554
409 ..... 409
1555 TGTCTGTGGCTTCACCTGTGAACCTGAGGCTCAAGACACACTACTCT 1604
410 ..... 417
1605 GGTGGTAAATGCTCAGAGCCTCCAGACGTCCTCCAGGCTGACGTGCC 1654
417 ..... 417
1655 AATGCAACAGACCCCTCACTCTATTCAATGTCAAGAAATGACGCAAG 1704
418 ..... 428
1705 AGCCTATGATGTGAATCCAGAACTCAGTGAAGCAACGCGAGTGACC 1754
429 GlyProPorphMeta1..... 434
1755 CAGTCACCTGGATGTCTCTATNGGCGCGACACCCCATCATTTCCCCC 1804
435 ..... 446
1805 CCAGACCTGCTTACTCTCGGAGCGAAC..... 1834
446 etValLeuSerThrLeuValSerGluAlaSerGlyHisProArgProThrIle 462
1835 ..... 1879
463 SerTrpAsnValAsnGlyThrAlaSerGluAspGlnAspProLys 479
1880 TCTTGGCGTATCAATGGATA..... 1908
479 gValLeuSerThrLeuAsnValLeu...ValTrpGluLeuLeuGluTr 495
1909 ACACACAGAGTCTCTTATGCGCAAAATACGCCAAATATATATACGGGA 1958
495 hrGlyValGluCysThrAlaSerAsn...AspLeuGlyLysAsnThrSer 510
1959 CC...TATGCCCTGTTTGTCTTACTGCTGCTACTGCGCGCATATATCC 2005
511 IleLeuPheLeuGluLeuValAsnLeuThrThrLeuThrProAspSerAs 527
2006 ATA..... 2037
527 nrThrThrGlyLeuSer 533
2038 AACTCTCTGCTCTCA 2056
seq_name: /cgnr2_6/prodata/2/ina/PC/US.COMB.seq:PCT-US93-00031-14
seq_documentation_block:
Sequence 14 Application PC/TUS9300031
GENERAL INFORMATION:
APPLICANT: Osborn, Laurelee
APPLICANT: Benjamin, Christopher D.
TITLE OF INVENTION: ANTIBODIES RECOGNIZING THE FOURTH
NUMBER OF SEQUENCES: 23
IMMUNOGLOBULIN-LINE DOMAIN OF VCAM1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Allegretti & Witcoff, Ltd.
STREET: 10 South Wacker Drive
CITY: Chicago
STATE: IL
COUNTRY: US
ZIP: 60606

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/00031
FILING DATE: 19930112
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: McNicholas, Janet M.
REGISTRATION NUMBER: 32,918
REFERENCE/DOCKET NUMBER: 92,306-A; D001 CIP PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 715-1000
TELEFAX: (312) 715-1234
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 2208 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2208
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 73..2208
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 1..72
PCT-US93-00031-14

alignment_scores:
Quality: 237.50 Length: 704
Ratio: 0.745 Gaps: 30
Percent Similarity: 45.312 Percent Identity: 20.739
alignment_block:
US-09-653-961-2 x PCT-US93-00031-14
Align seq 1/1 to: PCT-US93-00031-14 from: 1 to: 2208
33 ProGluLeuValGluVal.....GluValGlySe 42
||||| ||||| |||||
334 CCAGAACGGGTGAACTGACGCCCTCCCTTGGCAGCCAGTGAGCA 383
42 rThrAlaLeuLeuLysCysGlyLeuSerGlnSerGlnGlnGlyAsnLeuSerH 59
||||| ||||| |||||
384 GAACCTTACCCTACGCTCC.....CAGGTGGAGGGTGGGCGACCCC 424
59 IsValAspTrpPheSerValHisLysGluLysArgThrLeuIlePheArg 75
||||| ||||| |||||
425 GGGCGAAC.....CTCACCGTGGTGCCTCCGT 453
76 ValArgGlnGlyGlnGlyLysArgLysProGlyLysArgLysArgLysArg 92
||||| ||||| |||||
454 GGGAGAGAGCTCATGAGAGTCAAGAAATTCGTGAGAGATGCAGAGCA 503
92 uSerLeuGlnAspArgGlyAlaThrLeuAlaLeuThrGlnValThrProG 109
||||| ||||| |||||
504 CTCCTCGAAGAACCAAGATTTGGAGATACCTTACTCTGTGCATTTAGG 553
109 LAspArgLysGlyLeuLeuLysArgLysArgLysArgLysArgLysArg 120
||||| ||||| |||||
554 ATATTGGAAGATCTCTGTTGTCGAGGTAAATTTACACATATATATATG 603
121 .....ProArgSerGlnGlyLysArgLysArgLysArgLysArgLysArg 134
604 GATTCTGTGCCACAGTAAAGGAGGCTGTAAAGAAATGCAAGTCTACAT 653
134 sAlaProGluGluProAsnIleGlnValAsnProLeuGlyIleProVala 151

```

```
654 ATDCCAGAAATACAGTTATTTCTGTGAATCCA.....TCCACAA 694
      :::::::::::::: ::::::::::::::
151 snSerlysgluProgluGluValAlaThrCysValGlyLysAsnGlyTr 167
      :::::::::::::: ::::::::::::::
695 AGCTCCAAAGAGGTGCTGTGACCATGACCTGTCTCCAGCGAGATGA 744
      :::::::::::::: ::::::::::::::
168 ProIleProGluValIleTrpTrpLys.....AsnGlyArgProIle 181
      :::::::::::::: ::::::::::::::
745 CCAGCTCCAGAGATTTCTGGATAAGAAATTAAGATATGCG..... 786
      :::::::::::::: ::::::::::::::
181 uLysgluGluLysAsnArgValAlaIleGlnSerSerGlnThr..... 195
      :::::::::::::: ::::::::::::::
787 .....AATCTACAGCACCTTTCTGGAAATGACACCTTCACCT 823
      :::::::::::::: ::::::::::::::
196 .....ValGlnSerSerGlyLeuTrpTrpLeuGlnSerIle 207
      :::::::::::::: ::::::::::::::
824 TAATGCTATGAGGATGAGAAATCTGAAATTTATGTGTGTAAGAGTT 873
      :::::::::::::: ::::::::::::::
208 LeuLysAlaGlnLeuValLysGluAspLysAspAlaGlnPheTrpCysG 224
      :::::::::::::: ::::::::::::::
874 .....AATTGATT..... 882
      :::::::::::::: ::::::::::::::
224 uLeuAsnTrpArgLeuProSerGlyAsnHisMetLysGluSerArgGlu 241
      :::::::::::::: ::::::::::::::
883 .....GGGMAAAACAGAAAGAG..... 900
      :::::::::::::: ::::::::::::::
241 alThrValProValPheTrpProThrGluLysValTrpLeuGluValGlu 257
      :::::::::::::: ::::::::::::::
901 .....GTGGAAATTAATGTTCAAGAGAAACATTTACTGTGTGACATGCTC 945
      :::::::::::::: ::::::::::::::
258 ProValGlyMetLeu.....LysGluLysAspArgValGluLysCys 272
      :::::::::::::: ::::::::::::::
946 CCGAGACCCCGGATGCTGCTGCTGCTGAGTGAAGACATCACTGTGTGACATG 995
      :::::::::::::: ::::::::::::::
272 sLeuAlaAspGlyAsnProProHisPheSerIleSerLysGlnAsnPro 289
      :::::::::::::: ::::::::::::::
996 TAATGTCATGGCTGTGAATCCCATCTTTCTCTCGAGAACCCAGATAG 1045
      :::::::::::::: ::::::::::::::
289 roSerThrArgGluAlaGluGluGluTrpThrAsnAspArgGlyValLeu 305
      :::::::::::::: ::::::::::::::
1046 ACAGCCCTCTGACCGGAGAGTGAAGAGGAGGACCAATTCACAGCTG 1095
      :::::::::::::: ::::::::::::::
306 ValLeuGluProAlaArgLysGluHisSerGlyArgTrpGluCys.... 320
      :::::::::::::: ::::::::::::::
1096 ACCCTGAGCCCTGTGAGTTTGTGAGACGACACTTATCTGTGCAAGCT 1145
      :::::::::::::: ::::::::::::::
321 .....GlnGlyLeuAspLeuAspThrMetIleSerLeuLeu 333
      :::::::::::::: ::::::::::::::
1146 GACTTGTGACATTAAGAACTGGAAAGGAGAAATCCAGGTGACACTCTACT 1195
      :::::::::::::: ::::::::::::::
333 eArgLupProGlnLupLeuLeuValAsnTrpValSerAspValArgValSer 349
      :::::::::::::: ::::::::::::::
1196 CATTCCTCTAGAGAT..... 1209
      :::::::::::::: ::::::::::::::
350 ProAlaAlaProGluArgGln.....GlnGlySerSer 360
      :::::::::::::: ::::::::::::::
1210 .....CCGAAATCGAGATGAGTGTGCTGCTGCTGATGAGGAGCTC 1250
      :::::::::::::: ::::::::::::::
360 rLeuThrLeuThrCysGluAlaGlnSer.....SerGlnAspLeuG 374
      :::::::::::::: ::::::::::::::
1251 TGTCACTGTAGCTGCAAGGTTCTTACGCTTACCCCTTGACCGGCTGTG 1300
      :::::::::::::: ::::::::::::::
374 lupPheGluTrpLeuArgGluGluTrpGlyGlnValLeuGluArgGlyPro 390
      :::::::::::::: ::::::::::::::
1301 AGATGAAATTAATTAAGCGGAGACT.....ATTCTGAGAAATATAGAG 1344
      :::::::::::::: ::::::::::::::
391 ValLeuGlnLeuHisAspLeuLys..... 398
      :::::::::::::: ::::::::::::::
1345 TTTTGTGAGAAATACGATATGAAATCTTAGAGAAACAAAGTTTGGAAAT 1394
      :::::::::::::: ::::::::::::::
399 .....ArgGluAlaGlyGlyGlyTrpArgCysValAla 409
      :::::::::::::: ::::::::::::::

1395 GACCTTCATCCCTACCATTAAGATACATGGAAGAGCTCTTGTGTGACG 1444
      :::::::::::::: ::::::::::::::
409 lAserVal.....ProSerIleProGlyLeu 417
      :::::::::::::: ::::::::::::::
1445 CTAAAGTTACATTTTGAATGACATGAAATTCGAAACCAAGCAAGAGAGT 1494
      :::::::::::::: ::::::::::::::
418 AsnArgThrGlnLeuValAlaValAlaIlePheGlyProProTrpMetAl 434
      :::::::::::::: ::::::::::::::
1495 AGCGAAACACTTAATATGTCAAATGTTGCG.....CCC..... 1524
      :::::::::::::: ::::::::::::::
434 aPheLysGluArgLysValTrpValLysGluAsnMetValLeu..... 448
      :::::::::::::: ::::::::::::::
1525 ....AGAGTACAAACGCTCTGTGTCAGCCCTTCCTCCATCTGTGAGAGAG 1570
      :::::::::::::: ::::::::::::::
449 .....AsnLeuSerCysGluAlaSerGlyHisProArgProThr 461
      :::::::::::::: ::::::::::::::
1571 GCAGTTCTGTGAAATATGACATGCTTGAGCGAGGCTTTCCTGTGCGAAA 1620
      :::::::::::::: ::::::::::::::
462 lIleSerTrpAsnValAsnGlyThrAlaSerGluGlnAspGlnAspProGl 478
      :::::::::::::: ::::::::::::::
1621 ATCCTGTGAGC.....AGCAGCTCCCTAA 1646
      :::::::::::::: ::::::::::::::
478 nArgValLeuSerThrLeuAsnValLeuValThrProGluLeuLeuGluTr 495
      :::::::::::::: ::::::::::::::
1647 CGGGAGCTACAGCCTCTTCTGAAATGCAACTCCATCTTAATTTCTA 1696
      :::::::::::::: ::::::::::::::
495 hArgLysAlaGlu.....CysThrAlaSerAsnAspLeu 505
      :::::::::::::: ::::::::::::::
1697 CAATAATGCAAGATTCGCGGCTTATTATTATGGAAGAAATTAACAGGCT 1746
      :::::::::::::: ::::::::::::::
506 GlyLysAsnThrSerIleLeuPheLeuGluLeu..... 516
      :::::::::::::: ::::::::::::::
1747 GGAAGAGCAGAAAGAAAGATGAAATTAATTAATCAAGTACTGCCAAAGA 1796
      :::::::::::::: ::::::::::::::
517 ValAsnLeuThrThrLeuThrProAspSer.....AsnTrpTr 529
      :::::::::::::: ::::::::::::::
1797 CATTAACCTTAACGTTTCTCTGAGAGGTGTCAGAAAGAGAGACAGCTG 1846
      :::::::::::::: ::::::::::::::
529 hThrGlyLeuSerThrSerThrAlaSerProHisThr..... 541
      :::::::::::::: ::::::::::::::
1847 TCATCATCTCTTGTACATGTGAAATGTTCCAGAAACATGATAATCTCG 1896
      :::::::::::::: ::::::::::::::
542 .....ArgAlaAsnSerThrSerThrGluArgLys..... 551
      :::::::::::::: ::::::::::::::
1897 AAGAAAHAGCGAGACGAGGACGACGACTACTAAATCTATAGATGGCG 1946
      :::::::::::::: ::::::::::::::
552 .....LeuProGluProG 556
      :::::::::::::: ::::::::::::::
1947 CTATACATCCGAAAGCCCAAGTTGAGAGATGCGGAGATATATAGATGTG 1996
      :::::::::::::: ::::::::::::::
556 luserArgGlyValVal..... 561
      :::::::::::::: ::::::::::::::
1997 AATCTAATAAACAAAGTTGCTCAATTAAGAGTTTAACACTTGATGTT 2046
      :::::::::::::: ::::::::::::::
562 .....IleValAla 564
      :::::::::::::: ::::::::::::::
2047 CAAGAGAGAGAAACAAAGACTATTTTCTCTGAGCTTTCGTGCTGT 2096
      :::::::::::::: ::::::::::::::
564 aValIleValCysIleLeuValLeuAlaValLeuGlyAlaValLeuTrp 581
      :::::::::::::: ::::::::::::::
2097 CTATTTTGCATCTCTCTTAATATACCTGCCATTGGAGATGATATTAAT 2146
      :::::::::::::: ::::::::::::::
581 hLeuTrpTrpLys.....LysGlyLysLeuProCysAspArgSerGly 594
      :::::::::::::: ::::::::::::::
2147 TTGCAAGAAAGAACCAACATGAGAGGGGTCAATVAGCTTGTAGAAAGCAG 2196
      :::::::::::::: ::::::::::::::
595 LysGlnGluIle 598
      :::::::::::::: ::::::::::::::
2197 AAATCAAAAGTG 2208
```



```

68  ulysargthrleuilepheargvalarglnglnglnserglup 85
   |||::|||
203  GAAGCGAGCTCATCTTCCGTGTGGCCAGGGCCAGGGCCAGACGAGAC 252
   |||
85  rogllygluylrglulglnargleuSerleuGlnAspArglValaThrleu 101
   |||
253  CTGGGAGTAGAGAGAGCGGCTCAGCTCTCCAGACAGAGGGGCTACTCTG 302
   |||
102  AlaleuthGlnValThrProGlnAspGluThrleuLeuGlnGln 118
   |||
303  GCCCTGACTCAAGTCAACCCGCCCAAGAGCGGCATCTTCTTGCCAGGG 352
   |||
118  ylysarproargserglnGluThrArgTleGlnleuargValTyrlysa 135
   |||
353  CAAGCGCCCTGGGTCCCGAGAGTACCGCATCCAGCTCCGCGCTCAAG 402
   |||
135  lAProGlnGluProAsnleGlnValAsnProleuGlyTleProValAsn 151
   |||
403  CTCGGAGAGAGA .AACATCCAGGTCAACCCCTGGGCTCCCTGTGAAAC 451
   |||
152  SerlysgluProGlnGluValaThrCysValGlyArgAsnGlyTyrPr 168
   |||
452  AGTAAGAGACCTGAGAGAGGTGCTGCTAGGAGAGAACGGGTACCC 501
   |||
168  oileProGlnValleTrrPtyrlysaAsnGlyArgProleuLysgluGlu 185
   |||
502  CATCTCTCAAGTCATCTGTACAGAAATGCCGCTCGAAGAGAGAGA 551
   |||
185  ysAsnArgValHisTleGlnSerSerGlnThrValGluSerGlyleu 201
   |||
552  AGAACGGGTCCACATTCAGTCGTCCCAAGACTGTGAGTGCAGATGGTTTG 601
   |||
202  TyrThrleuGlnSerTleleuysAlaGlnleuVallyGlnAspLyss 218
   |||
602  TACACCTTCAGAGTATCTTGAAGGCACAGCTGTGTAAGAGACAAAGA 651
   |||
218  palaglnPheTyrCysGluLeuAsnTyrArgleuProserGlyAsnHis 235
   |||
652  TGCCCGAGTTTACTGTGAGCTCACTACAGCGGCTGCCCATGGAAACACA 701
   |||
235  et. LysgluSerArgGluValaThrValProValPheTyrProThrGly 251
   |||
702  TGAAAGAGTCCAGGAGAGTCAACCCGCTGTTTCTTACCACCAAGAAA 751
   |||
251  sValTlePleuGlnValGluProValGlyMetleuLysgluGlyAspArgV 268
   |||
752  AGTGTGGCTGAGAGTGAAGCCCGTGGAAATGCTGAAGAGAGGAGCCGC 801
   |||
268  alGluThrArgCysleuAlaAspGlyAsnProProThrHisPheSerile 284
   |||
802  TGGAAATCAAGGTGTTGGCTGATGGACGCCCTCCACACACTTCACATC 851
   |||
285  SerlysglnAsnProSerThrArgGluAlaGluGluGluThrThrAsn 301
   |||
852  AGCAAGCAGAAACCCAGACACAGAGGAGCAGAGAGAGACAAAC .AACCA 900
   |||
301  paenglyValleuValleuGlu...ProAlaArgLysglnHisSerGlyA 317
   |||
901  CAACGGGGTCTCTGCTGAGAGTGGCCGGAAGAACACATGGGSG... 946
   |||
317  rgtTyrGluCysGlnGlyleuAspPleuAspThrMetIleSerleuLeuSer 333
   |||
947  ..TATGAAATGTCAGGSGCTG...ACTGAAACATGATATATCTST 985
   |||
334  GluProGlnGluLeuValAsnTyrVal.SerAspValArgValSerP 350
   |||
986  GAGTGAACACAGGAMTATGTGTACTATGTTTATM.....GTGGATK 1026
   |||
350  roAlaIaIaProGluArgGlnGluGlySerSerleuThrleuThrCysGlu 366
   |||
1027  ATCCGGCGCCCTGAGAAAGAGACACAGCTCAGCTGACTGTGAGCAAAATA 1076

```

```

367  AlaGluSerSergin 371
   |||::|||
1077  GCAGACTCGATTGCG 1091
   |||
seq_name: gb_est1:AL556415
seq_documentation_block:
LOCUS      AL556415                      956 bp    mRNA       linear    EST 16-FEB-2001
DEFINITION AL556415 LRT_NFL006.PL2 Homo sapiens cDNA clone CS0DK004YG06 5
ACCESSION  AL556415
VERSION     AL556415.1  GI:12899069
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 956)
AUTHORS    Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL    Unpublished (2001)
COMMENT     Contact: Genoscope
             Genoscope - Centre National de Sequencage
             BP 191 91006 Evry cedex - France
             Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
     source
         1..956
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="CS0DK004YG06"
             /clone_lib="LRT_NFL006.PL2"
             /tissue_type="placenta"
             /note="Vector: PCWVSPORT 6; Site_1: NotI; 1st strand cDNA
             was primed with a NotI-oligo(dT) primer. Five prime end
             enriched, double-stranded cDNA was digested with Not I and
             cloned into the Not I and Eco RV sites of the PCWVSPORT 6
             vector. Library was normalized. Library was constructed by
             Life Technologies. Contact : Feng Liang Life Technologies,
             Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
             Email : fliang@lifetech.com URL :
             http://fulllength.invitrogen.com"

BASE COUNT      214 a      272 c      292 g      171 t      7 others
ORIGIN
alignment_scores:
    Quality: 1556.00      Length: 320
    Ratio: 4.928          Gaps: 0
    Percent Similarity: 98.750      Percent Identity: 97.812

alignment_block:
US-09-653-961-2 x AL556415 ..
Align seg 1/1 to: AL556415 from: 1 to: 956
1 MetGlyleuProArgleuValCysAlaPheleuLeuAlaIaCysCysCy 17
1 ATGGGGCTTCCAGAGCTGTGCTGGCCCTTCTGCTCGCGGCTCTCTG 50
17 scyproargvalAlaGluValArglySerThrAlaLeuLeuLyCysGlyleu 50
51 CTGTCTCGCGGTGGGGGTGTGCCGGAGAGCTGAGAGCGCTCGGCTG 100
34 luleuValGluValGluValGlySerThrAlaLeuLeuLyCysGlyleu 50
101 AKCTGTGTGAGGTGAATKGACAGACAGCCCTTGAATKGAGGCTC 150
51 SerGlnSerGlnGlyAsnLeuSerHisValAspTrrPheSerValHisly 67
151 TCCAGTCCCAAGGCACCTCAGCCATGTGAGCTGTGTTCTGTTCACAA 200
67 sgluysargthrleuilepheargvalarglnglnglnserg 84

```

```

|||||
201 GGAGAAAGGAGCGCATCTTCCTGTCGCGCCAGGCGCCAGCGCCAGACG 250
84 LIPROGLYGLUTYRGLUGLInArgLenseLInAspArgLAlaThr 100
251 AACCTGGAGACGACGAGCGGCTGCTCCCTCCAGACGAGGCGGCTACT 300
101 LeuAlaLeuThrGlnValThrProGlnAspGluArgLLeuLeuGly 117
301 CTGGCCCTGACTCAAGTACACCCGCCAAGAGACGACGATCTTCTGTCCA 350
117 nGlyLysArgProArgSerGlnGlnGlyArgLLeuLeuArgValTyrL 134
351 GGCGCAAGGCGCCCTCGTCCAGAGAGTACGACATCCAGCTCCGCTCTACA 400
134 yAlaLProGlnGluProAsnLLeuValAspProLLeuGlyLeuProVal 150
401 AAGCTCCGAGAGG. GCAAAATCCAGSTCAACCCCTGGGCTCCCTGTX 449
151 AsnSerLysGluProGlnGluValAlaThrCysValGlyArgAsnGlyTyr 167
450 AACAGTAAGAGACCTGAGAGAGTGGCTACCTGTAGAGGAGGAGGAGTA 499
167 rProLLeuProGlnValLeuTyrTyrLysAsnGlyArgProLeuLysGln 184
500 CCCCATCTCTCAAGTATCTGTGTCAAGATGGCCGCTCTGAGAGGAGG 549
184 LulysAsnArgValHisLLeuLLeuSerSerGlnThrValGluSerSerL 200
550 AGAAGACCGGGTCCACATTCAGTCCGACAGCTGTGAGTGCAGTGTGT 599
201 LeuTyrThrLeuGlnSerLLeuLeuValGlnLeuValLysGlnAspL 217
600 TTGTACCTTGCAGAGTATCTGAGGACACAGCTGTGTAAAGAGACAA 649
217 sAspAlaGlnPheTyrCysGluLeuAsnTyrArgLeuProSerGlyAsnH 234
650 AGATGCCAGATTACTGTGAGCACTACACGCTGCCAGTGGGAAAC 699
234 iMetLysGlnSerArgLValAlaThrValProValPheTyrProThrGlu 250
700 ACATGAGAGAGTCCAGGAACTACGCTCTTTTACCCGACAGAA 749
251 LysValThrLeuGlnValAlaProValGlyMetLeuLysGlnLysAsp 267
750 AAAGTGTGCTGAGAGTGGAGCCCTGTGGAACTGAGAGAGAGGAGACG 799
267 gValGlnLeuArgCysLeuAlaAspGlyAsnProProOHISpSerL 284
800 CGTGGAAATCAGGTGTGTGCTGATGCAACCTCCACCACTTCAACA 849
284 LeuSerLysGlnAsnProSerThrArgLLeuLLeuGlnGlnLLeuThrAsn 300
850 TCARCAAGCAGAACCCAGC. ACCAGGAGGACAGAGAGAGAGAAC 897
301 AspAsnGlyValLeuValLeuGlnProAlaArgLysGlnHisSerGly 317
898 GACAAAGGAGGCTCTGTGCTGAGACCTGCCGGA. GAAACACAGTGGCG 946
317 gTyrGlnCys 320
947 CTATGATATGT 956
seq_name: gb_est1:AL541635
seq_documentation_block:
LOCUS AL541635 948 bp mRNA linear EST 16-FEB-2001
DEFINITION AL541635 LTI-FL002_P11 Homo sapiens cDNA clone CS0DE003YA16 5 prime
, mRNA sequence.
ACCESSION AL541635
VERSION AL541635.1 GI:12872890
KEYWORDS EST.
SOURCE human.

```

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 948)
AUTHORS L.I.W.B., Gruber,C., Jesssee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1..948
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DE003YA16"
/location="Lib-FL002_P11"
/label="DH10B"
/note="Organ: Placenta; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a Molt-Oligo(dT) primer. Five prime
end and enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
fliang@lifetechn.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 213 a 271 c 296 g 166 t . 2 others
ORIGIN
alignment_scores:
Quality: 1528.00 Length: 314
Ratio: 4.945 Gaps: 2
Percent Similarity: 98.408 Percent Identity: 98.408
alignment_block:
US-09-653-961-2 x AL541635
Align seg 1/1 to: AL541635 from: 1 to: 948
1 MetGlyLeuProArgLeuValCysAlaPheLeuLeuAlaLysCysGly 17
|||||
12 ATGGGGCTTCCAGGCTGTGTGCTGTGTGTGCTGCGCGCTGTGCTG 61
17 scysProArgValAlaGlyValProGlyGlnAlaGlnGlnProAlaProG 34
|||||
62 CTGTCTCGCGTGGGCTGTGCTGCGGAGGCTGTGAGCAGCTGTGCTG 111
34 LuleuValGlnValGlnValGlySerThrAlaLeuLeuLysCysGlyLeu 50
|||||
112 ACCGTGTGAGGTGGAGAGTGGGACAGACACCTTGTGAAAGTGGGCTC 161
51 SerGlnSerGlnGlyAsnLeuSerHisValAspTyrPheSerValHisL 67
|||||
162 TCCCACTCCCAAGGCAACCTCAGCATGTGATGCTTTTCTGTGCAACA 211
67 sGlyLysArgThrLeuLLeuLLeuPheArgValArgGlnGlnGlnLysArg 84
|||||
212 GGAGAAAGCGGAGCGTCATCTTCGTGTGCGCCAGGCGCAGGCGCAGACG 261
84 LueProGlyLLeuTyrGlnGlnArgLeuSerLLeuGlnAspArgLLeuThr 100
|||||
262 AACCTGGAGATACGACGAGGCTCAGCTCCAGACAGAGAGGAGCTACT 311
101 LeuAlaLeuThrGlnValThrProGlnAspGluArgLLeuLeuGly 117
|||||
312 CTGGCCCTGACTCAAGTACACCCGCCAAGAGACGACGATCTTCTGTCCA 361
117 nGlyLysArgProArgSerGlnGlnGlyArgLLeuLeuArgValTyrL 134
|||||
362 GGCGCAAGGCGCCCTCGTCCAGAGAGTACGACATCCAGCTCCGCTCTACA 411

```

134 ysalaprogliupprokaniieglnvalasproleugliyleproval 150
 412 AAGTCCGGAGAGANCA. AACATCCAGGTCAACCCCTGGGCATCCCTGTG 460
 151 Aasnerlysgliuprogliuvalalathrcysvalglarsanglyty 167
 461 AACGTAAGAGAGCTGAGAGGTGCTACCTGCTGAGGAGAGAGGTA 510
 167 rprolietroglnvaliiletrpyrlysaanglyarproleuylsglug 184
 511 CCCCATCCCTCAAGTCACTCTGCTACCAAGATGGCCCGCTCGAAGAGAG 560
 184 lulyasnarargvalhisileginserserlgnthralgluserserly 200
 561 AGAAGAACCGGTCACATTCAGTCTGTCAGACAGCTGGAGTGCAGTGT 610
 201 leuylthreuglnserileuylsalaglnleuvalylsgliuaply 217
 611 TTGTACACCTTCAGAGATTTCTTCAGACGACACCTGCTTAAAGAACAG 660
 217 sasralaglnpheylrcysgluleuasnryargleuproserylasn 234
 661 AGATGCCAGTCTTACTGAGCTCAACTACCGGCTGCCAGTGGAGACC 710
 711 ACATGAAGAGAGTCCAGGAGAGTCAACGCTCCTGTTTCTACCCACAG 760
 250 ulysvaltrpleuglnvalgluprovalglymetleuylsgliuaply 267
 761 AAAAGTGTGCTGAGAGTGGAGCCGCTGGATGCTGAAGAGAGAGAG 809
 267 rglvalglulearagcysleualaspglyasnproprothiasheer 283
 810 C.GTGGAATCAAGTGTGTTGGCTGATGCA. CTTCCACACACTTCAC 858
 284 lleserlysgliuasnpserthrarglualagluglunththas 300
 859 ATCCAGACAGACACCCAGCACAGGAGAGAGAGAGAGACACACAA 908
 300 naspsan.glyvalleuvalleuglnproalaarglys 312
 909 CGACACAGGCTGCTGCTGCTGAGAGCTGCCGAGAG 946
 seq_name: gb_est1:AL554839
 seq_documentation_block:
 LOCUS AL554839 1022 bp mRNA linear EST_16-FEB-2001
 DEFINITION AL554839 LTI_NFL006.PL2 Homo sapiens cDNA clone CS0D1087Y10 5
 prime, mRNA sequence.
 ACCESSION AL554839
 VERSION AL554839.1 GI:12896003
 KEYWORDS
 EST.
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 FEATURES
 source
 1..1022
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0D1087Y10"
 /clone_lib="LTI_NFL006.PL2"
 /tissue_type="Placenta"

/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA
 was primed with a NOTI oligo(dT) primer. Five prime end
 enriched double-stranded cDNA was digested with Not I and
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6
 vector. Library was normalized. Library was constructed by
 Life Technologies. Contact : Feng Liang Life Technologies,
 a division of Invitrogen 9800 Medical Center Drive
 Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
 Email : fliang@life.com
 http://fulllength.invitrogen.com"

BASE COUNT 237 a 306 c 289 g 184 t 6 others
 ORIGIN

alignment_scores:
 Quality: 1520.00 Length: 297
 Ratio: 5.118 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 99.327

alignment_block:
 US-09-653-961-2 x AL554839 ..
 Align seg 1/1 to: AL554839 from: 1 to: 1022

350 PROALAAPROGLUARGGLUGLIGLYSERSERLEUTHRCYSGI 366.
 2 CCCGACACCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 51
 366 UALAGLUSERSERGLINASPLEUGLUPHEGINTRPLEUARGGLUGLUTHRG 383
 52 GCGAG 101
 383 LYGlnvalleuglnuarggluprovalleuaglinsasplespulyarg 399
 102 GCCAGTGTGCGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 131
 400 GUALAGLIGLYGTYRARGCYSAVALASERVALPROSERLIEPROGL 416
 152 GAGGACAG 201
 416 YLEUASNARGTHRGlnleuvalasvalalalepheglyproptfrm 433
 202 CCTCAACCCGACAGAGCTGCTCAACGCTGGCCATTTTGGCCCCCTTGA 251
 433 etAlapheuglnuargglusvaltrpyvalysgluasnmetvalleuasn 449
 252 TGGCATTCAG 301
 450 leusercysglualaserglyhisproargprothrilesertpasaava 466
 302 CTGCTTGTGAAGCGTCAGGGCACCCCGGCGCACCATCTCTCTGGAACGT 351
 466 laseglythralsasergluglnaspglnaspproglnatryalileusert 483
 352 CACGGCAGGAG 401
 483 hrleuasnvalleuvalthrprogluleuuglnthrylyvalglucys 499
 402 CCCGATATGCTCTCTGACCCCGAGAGCTGTGGAGCAGAGGTGAATGC 451
 500 ThrAlaserasnasplesuglylysaasnThrserileuapheluglule 516
 452 ACGGCTCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 501
 516 uvalasleuvalthrleuvalthrproaspsersanrthrthrcglyleus 533
 502 GGTCAATTTTACACCTTCACACAGAGAGAGAGAGAGAGAGAGAGAGAG 551
 533 erThrserThrAlasSerProHISrArgAlaAsnSerThrThGlu 549
 552 GCACCTTCACCTCGAGTCTCATATCCAGAGAGAGAGAGAGAGAGAGAGAG 601
 550 Arglyleuprogliuprogliuserarglyvalvaliilevalalavalaii 566

```

|||||
602 AGAAGCTCCGCGAGCGGAGAGCCGGGCGTGTATCTGCGCTGAT 651
566 eValcyslleuValleuAlaValleuGlyAlaValleuTyrrpheLeuT 583
652 TGtGtCACTGCTGCTGCGGCGTGTGCGCTGCTCTATTTCTCT 701
583 ytllysGlylyLeuProCysArgArgSerGlylySglnIleThr 599
702 ATAGAAGGGCAAGCTGCGGTCAGGCGCTCAGGAGAGCAGATCAAG 751
600 leuProSerArglySerGlnIleuValIleValIleValIleSerAsp 616
752 CTACCCCGCTCTCGTAAGAGCGAAGCTGTGAGTTGAAGTAAGTAATA 801
616 sIeuProGlnIleuMetGlyLeuLeuGlnIleuGlnIleuSerGlyAsp 633
802 GCTCCAGAGAGATGGGCTCTGCGAGGCGACACGCGGTGACAGAGG 851
633 lAbroGlyaspGlnIleuGlylyTyrrlleaspLeuArgHis 646
852 CTCGGGAGACAGGGGAGAAATACATCATCTGAGGCAT 892

seq_name: 9b_est1:AL545925
seq_documentation_block:
LOCUS AL545925 879 bp mRNA linear EST 16-FEB-2001
DEFINITION AL545925 l7l_nfl006.pl2 Homo sapiens cDNA clone CSDD1023YF23 5
prime, mRNA sequence.
ACCESSION AL545925
VERSION AL545925.1 GI:12878563
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 879)
AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1..879
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSDD1023YF23"
/clone_1lb="l7l_nfl006.pl2"
/issue_type="Placenta"
/note="Vector: pCMVSPORT 6. Site 1: NotI. 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifestech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 226 a 216 c 276 g 157 t 4 others
ORIGIN
alignment_scores:
Quality: 1480.50 Length: 293
Ratio: 5.158 Gaps: 2
Percent Similarity: 97.952 Percent Identity: 97.611
alignment_block:
US-09-653-961-2 x AL545925

```

Align seg 1/1 to: AL545925 from: 1 to: 879

```

153 lysGluProGlnIleuValAlaThrCysValIleGlyArgAsnGlyTyrrProI 169
2 MAGAGACCTGAGAGAGTGGCTACTGCTGTGTAGGAGAGAACGGGTACCCAT 51
169 eProGlnValIleTyrrlyAsnGlyArgProLeuTyrsGlnIleuTyra 186
52 TCCTCAAGTCACTGTGTCAAGAAATGGCGGCTCTGAAGAGGAGGAAGA 101
186 snArgValHisIleGlnSerSerGlnThrValGlnSerSerGlyLeuTyrr 202
102 ACCGGGTCCACATTCATGCTCCAGACTGTGAGTGCAGTGTGTGTATC 151
203 ThrLeuGlnSerIleLeuTyAlaGlnLeuValIleGlnAspIleAspAl 219
152 ACCTGCAGATGATTCGTAAGGCACAGCTGTGTAAGAAAGACAAAGATGC 201
219 eGlnPheTyrrCysGlnLeuAsnTyrrArgLeuProSerGlyAsnHisMet 236
202 CCAGTTTACTGTGTAGCTACACTACCGGCTGCCACTGGAAACCATGA 251
236 ysgIuSerArgGluValIleThrValProValPheTyrrProThiGlnVal 252
252 AGGAGTCCAGGGAAGTCCGCTCCCTGTCTTCTACCCGACAAATAAAGTG 301
253 TrpLeuGlnValIleuProValIleuLeuTyrsGlnIleuAspArgValG 269
302 TGCTCTGAAGTGGAGCCCGTGGAAATCTGAAGAAAGGGGAGCGGTGCA 351
269 uIleArgCysLeuAlaAspGlyAsnProProPheHisPheSerIleSer 286
352 AATCAGGTGTGTGGCTGATGGCAACCTCCACACACTTCACATCAGCA 401
286 ysgIuAsnProSerThrArgGlnIleuGlnIleuThrThrAsnAsp 302
402 AGCAGAACCCCGACACAGGAGGCAAGAGAAACCAACACACACAC 451
303 GlyValleuValleuGluProAlaArgIleGlnIleSerGlyArgTyrr 319
452 GGGGTCTGTGTCTGTGAGCCCTGCGGAGAAAGACAGATGGGCTATGA 501
319 uCysGlnIleuAspLeuAspThrMetIleSerIleuLeuSerGluPro 336
502 ATGTCAAGGCGCTGGACTTGGACACCATGATATGCTGAGTGAACCC 551
336 IncIleuLeuValAsnTyrrValSerAspValArgValSerProAla 352
552 AGGAATACTGTTGAATATGTGTGACGTCCGAGTGAAGTCCCGCACCC 601
353 ProGlnArgGlnIleuGlySerSerLeuThrLeuThrCysGlnAlaGln 369
602 CCGAAGAGACAGAGAGCGAGCGCTCACCTGACTGTGAGGACAGAG 651
369 rSerGlnAspLeuGlnPheGlnTrpLeuArgGlnIleuThrGlyGlnVal 386
652 TACCCAGGACCTCGAGTTCAGTGTGCTGAGARARARARAGAGGCGAGG 701
386 euGlnArgGlyProValleuGlnIleuHisAspLeuTyrsArgGlnIle 402
702 TGGAAGAGGGGCGCTGCTTCAAGTTGATGACCTGAAGAGGAGAGAG 751
403 GlyGlyTyrrArgCysValAlaSerValProSerIleProGlyLeuAsn 419
752 GCGGCTATCGCTGCGTGGCGTGTGCTCCAGCATACCGGCGCTGAACG 801
419 gThrGlnIleuValAsnValAlaIlePheGlyProThrPheMetIlePhe 435
802 CACACAGCTGTGCACAGCTGCA...TTTTTGCCCGCTTGATGATGCATTC 848
436 LysGluArgIleValIleValIleGln 444
849 AAGGAAAGAGAGGTGTGGTAAAAAA 875

```

```

seq_name: gb_est1:AL548612
seq_documentation_block:
LOCUS AL548612 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0D1036YB01 5
DEFINITION prime mRNA sequence.
ACCESSION AL548612
VERSION AL548612.1 GI:12883791
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 868)
AUTHORS L.L.W.B., Gruber,C., Jesssee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr.

FEATURES
source
1..868
/organism="Homo sapiens"
/txid="taxon:9606"
/clone="CS0D1036YB01"
/clone_lib="LTI_NFL006_PL2"
/lisue_type="Placenta"
/note="Vector: pCMVSPORT 6; site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifestech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 188 a 249 c 269 g 160 t 2 others

Alignment_scores:
Quality: 1426.50 Length: 288
Ratio: 5.023 Gaps: 2
Percent Similarity: 98.611 Percent Identity: 98.611

Alignment_block:
US-09-653-961-2 x AL548612 ..
Align seg 1/1 to: AL548612 from: 1 to: 868

```

```

84 1uproglyglntrglnlgnlnarlglsuerrleuGlnAAspArgGlyAlaThr 100
|||||
262 AACCTGGGGAGTACGACGACGCGCTCGCTCCAGACAGAGGGCTACT 311
|||||
101 LeuAlaLeuThrGlnValThrProGlnAspGlnAlaGlyIlePheLeuGly 117
|||||
312 CTGGCCCTGACTCAAGTCACTCCCAAGAGAGAGCGCATCTTCTGTCCCA 361
|||||
117 nglylsargProargSerGlnGlntrglntrglnleuAlaArgValTyr 134
|||||
362 GGGCAAGCCCTCGCTCCAGAGTACCGCATCCAGCTCCGCTGACCA 411
|||||
134 ysaIapProGlu.GluProasnIleGlnValaAsnProLeuGlyIleProVa 150
|||||
412 AACCTCGGAGGSCA...ANCATCCAGGTCAACCCCTGGGCGATCCCTGT 458
|||||
150 IasnSerLysGluProGlnGlnValAlaThrCysValGlnArgAsnGly 167
|||||
459 GAACAGTAAAGAGCCCTGAGAGGTCCCTACCTGTAGAGAGAACGGT 508
|||||
167 yrrProIleProGlnValIleThrTyrLysAsnGlyArgProLeuGly 183
|||||
509 ACCCATCTCTCAAGTCACTGGTACAGAAATGCGCGCTCTGAAGAG 558
|||||
184 GluLysAsnArgValHisIleGlnSerSerGlnThrValGlnSerSer 200
|||||
559 GAGAAAGACCGGCTCCACATTCATGCTCCAGACTGTGAGTGAAGTGG 608
|||||
200 yLeuTyrThrIleGlnSerIleLeuLysAlaGlnLeuValLysGlnAsp 217
|||||
609 TTGTGTACACTTCGACAGATATTCGAAGACACAGCTGTAAAGAAACA 658
|||||
217 ysaAspAlaGlnPheTyrCysGlnLeuAsnTyrArgLeuProSerGlyAsn 233
|||||
659 AAGATCCCATTTTACGTAGAGTCACTACCGGCTCCAGAGGAGAA 707
|||||
234 HisMetLysGlnSerArgGlnValThrValProValPheTyrProThr 250
|||||
708 CACATGAAGAGACTCCAGAGAGTACCGCTCTGTTCACCGAGACA 757
|||||
250 ulysValTrpLeuGlnValGluProValGlyMetLeuLysGlnLysAsp 267
|||||
758 AAAAGTGTGCTGAGAGTGAAGCCCTGGGAATGCTGAAGAGAGGAGCA 806
|||||
267 rGValGluIleArgCysLeuAlaAspGlyAsnProProThrHisPheSer 283
|||||
807 C.GTGCNAATCAGGTGTTGGCTGATGACCAACCTCCACCACTTACG 856
|||||
284 IleSerLysGln 287
|||||
857 ATCAGCAACAG 868

```

seq_name: gb_est1:AL569913

seq_documentation_block:

LOCUS AL569913 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0D1001YH03 5

DEFINITION prime mRNA sequence.

ACCESSION AL569913

VERSION AL569913.1 GI:12925725

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

1 (bases 1 to 1019)

AUTHORS L.L.W.B., Gruber,C., Jesssee,J. and Polayes,D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr.

FEATURES
SOURCE

Location/Qualifiers
1. .1019
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS01016YD21"
/issue_lib="PL1.NFL006.PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by life technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 243 a 292 c 291 g 189 t 4 others

ORIGIN

alignment_scores:
Quality: 1424.00 Length: 281
Ratio: 5.122 Gaps: 0
Percent Similarity: 98.932 Percent Identity: 98.577

alignment_block:
05-09-653-961-2 x AL569913 ..

Align seg 1/1 to: AL569913 from: 1 to: 1019

```

366 gluAlaGluSerGlnAspLeuGluPheGlnTrpLeuArgGluLth 382
|||||
44 GAATTCGGGATGTCAGACCTGAGTCCAGTGGCTGAGAGAGAGAG 93
|||||
382 TGLGlnValLeuGluArgGlyProValLeuGlnLeuHisAspLeuLys 399
|||||
94 AGCCGAGGTGCTGGAAGGGGCTGTGCTCAGTTGACAGACTGAAC 143
|||||
399 TGLValAlaGlyGlyTyrArgGlyValAlaSerValProSerLlePro 415
|||||
144 GGGAGGCGAGAGGGGCTGATGCTGGGGTGTGCGCAGATATCCC 193
|||||
416 GlyLeuAsnArgThrGlnLeuValAsnValAlaLlePheGlyProPro 432
|||||
194 GGCCTGAACCCGACACCTGCTCAACGTGGCAATTTTGCCCCCTTG 243
|||||
432 PheTAlaPheLysGluArgLysValTrpValLysGluAsnMetValLeu 449
|||||
244 GATGGCATTCAGAGAGAGAGAGTGTGGTCAAGAAATATGTGTGA 293
|||||
449 snLeuSerGlyGluAlaSerGlyHisProArgProThrLleSerTrpAsn 465
|||||
294 ATCTGCTGTGAGAGCTCAGAGCCGCCGCCCATCTCCGAGAAC 343
|||||
466 ValAsnGlyThrAlaSerGluGlnAspGlnAspProGlnArgValLeuSe 482
|||||
344 GTCAAGCGCAGCGCAAGTCAACAGCAAGATCCACAGCGAGTCTGAG 393
|||||
482 TrpThrLeuAsnValLeuValThrProGluLeuLeuGluThrGlyValGlu 499
|||||
394 CACCTGAATGTCCTCGTACCCCGAGAGCTGTGGAGACAGGTGTGAAAT 443
|||||
499 ysrThrAlaSerAsnAspLeuGlyLysAsnThrSerLleLeuPheLeuGlu 515
|||||
444 GCAGGGCTCCAGAGACCTGGGCAAAACACAGCATCTCTCTGGAG 493
|||||
516 LeuValAsnLeuThrThrLeuThrProAspSerAsnThrThrGlyLe 532
|||||
494 CTGTCAATTTTAACACCTCACACAGACTCCAAACACACACCTGGCCT 543
|||||
532 userThrSerThrAlaSerProHisThrArgAlaAsnSerThrSerThg 549
|||||

```

544 CAGCACTTCACACTGCGACCTCCTCATACAGAGCCAAACAGCACCCTCCACAG 593

549 luArgLysLeuProGluLeuProGluSerArgGlyValValLleValAlaVal 565

594 AGAGAAAGCTCCGGAGCGGAGAGCCGGGCGGTGCATCGTGGCTG 643

566 lleValLysLleLeuValLeuAlaValLeuGlyValAlaValLeuTyrPhe 582

644 ATTTGTGATCTCTGCTGCTGCGGTCTGCGGCTCTCTCTATTTCC 693

582 utyrLysLysGlyLysLeuProGlyArgArgSerGlyLysGlnLleLter 599

694 CTTATAGAGGCGCAGCTGCCGTGCGAGCGCTCAGAGGAGCAGAGATCA 743

599 hrLeuProProSerArgLysSerGluLeuValAlaGluValLysSerAsp 615

744 CGCTGCCCGCTGCTGTAGAGCGAATTTGATTTGAAGTTAACTGAGAT 793

616 LysLeuProGluGluMetGlyLeuLeuGlnLysSerSerGlyAspLysAr 632

794 AACCTCCAGAGAGAGATGGGCTCCTCGCAGGCGCAGCAGCGGTGACAGAG 843

632 galAProGlyAspGlnGlyLysTyrLleAspLeuArgHis 646

844 GCCTCCGGGAGACAGGAGAGAAATCATCATCTGAGGCAT 886

seq_name: gb_est1:AL548357

seq_documentation_block:
LOCUS AL548357 864 bp mRNA linear EST 16-FEB-2001
DEFINITION AL548357 LFI_NFL006.PL2 Homo sapiens cDNA clone CS01016YD21 5
prime, mRNA sequence.
ACCESSION AL548357
VERSION AL548357.1 GI:12883284
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 864)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope, Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
SOURCE

1. .864
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS01016YD21"
/issue_lib="PL1.NFL006.PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by life technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 187 a 249 c 265 g 160 t 3 others

ORIGIN

alignment_scores:
Quality: 1356.50 Length: 279
Ratio: 4.951 Gaps: 3
Percent Similarity: 98.208 Percent Identity: 98.208

alignment_block:

US-09-653-961-2 x AL548357 ..

Align seg 1/1 to: AL548357 from: 1 to: 864

```

1  MetGlyLeuProArgLeuValCysAlaPheLeuLeuAlaAlaCysCysCys 17
4  ATGGGGCTTCCTCCAGGCTGGTCTGCGCTTCTGCTGCGCGCTGCTGCTG 53
17  sCysProArgValAlaGlyValProGlyGluAlaGluGlnProAlaProG 34
54  CTGTCTCGCGCTGCGGGGTGTGCTCCCGAGAGCTGAGAGAGCTGCGCTG 103
34  LuLeuValGluValGluValGlySerThrAlaLeuLeuLysCysGlyLeu 50
104  AGCTGGTGGAGGTGGAAAGGGGACACAGACAGCCCTTCTGAAGTGGCGCTC 153
51  SerGlnSerGlnGlyAsnLeuSerHisValAspTrpPheSerValHisLys 67
154  TCCAGAGTCCCAAGGCAACCTCAGCCATGTGAGCTGGTTTCTGTCACAA 203
67  sGluLysArgThrLeuIlePheArgValArgGlnGlyGlnGlyGlnSerG 84
204  GGAGAAAGCGGAGGCTCATCTTCCGTGCGCGAGGCGGAGGCGGAGCG 233
84  LuProGlyGluThrGluGlnArgLeuSerLeu.GlnAspArgGlyAlaThr 100
254  AACCTGGGGAGAGTACGAGCAGCGGCTCAGCCCTACAGAGAGAGAGGCTAC 303
100  rLeuAlaLeuThrGlnValThrProGlnAspGlyArgIlePheLeuGly 117
304  TCTGGCCCTGACTCAGTCACACCCGCCAAGAGAGAGGCTTCTTGAGCC 353
117  LngLysArgProArgSerGlnGlyUtrArgIleGlnLeuArgValAlaTrp 133
354  AGGGCAAGCGCTGCTGGTCCAGAGATCCGATCCAGGCTCCGCTGCTAC 403
134  LysAlaLeuProGlu.GluProAsnIleGlnValAsnProLeuGlyIlePro 150
404  AACCTCCGAGGAGSCA...AACATCCAGGTCAACCCCTGGGCGATCCCTG 450
150  AlaAsnSerLysGluProGluGluValAlaThrCysValGlyArgAsnGly 166
451  TGAACATTAAGAGAGCTGAGAGAGGTGCTACCTGTAGGAGAGAAAGCG 500
167  TyrProIleProGlnValIleTrpTyrLysAsnGlyArgProLeuLysGln 183
501  TACCCCATTCCTCAAGTCACTGTGTAACAAGATGGCGGCTCTGAAGGA 550
183  uGluLysAsnArgValAlaHisIleGlnSerSerGlnThrValGluSerSer 200
551  GGGAGAAAGACCGGGTCCACATTCAGTCGTCACAGACTGGAGTGCAGT 600
200  1yLeuTyrThrLeuGlnSerIleLeuLysAlaGlnLeuValLysGlnAsp 216
601  GTTTCACACCTTTCAGAGATATTCTGAAGGACACAGCTGTAAAGAAAGAC 650
217  LysAspAlaGlnPheTyrCysGluLeuAsnTyrArgLeuProSerGlyAs 233
651  AAAGATGCCAGCTTACTGTAGCTCAACTAACCGGCTGCCCATGGGGA 699
233  nHisMetLysGluSerArgGluValAlaThrValProValPheTyrProThr 250
700  .CAACTGAAGAGAGTCCAGAGAGTCAACCGTCCGTGTTTTCACCCGAGAG 749
250  LuLeuValAlaTrpLeuGluValGluProValGlyMetLeuLysGluGlyAsp 266
750  AAAAAGTGTGGCTGGAAGTGGAGCCCGTGGGAATGCTGAAGAGAGGGGAC 799
267  ArgValGluIleArgCysLeuAlaAspGlyAsn 277
800  GC.GTGGAAATCAGGTGTGCTGATGATGCAAC 831

```

seq_name: gb_est1:AU119585

seq_documentation_block:

LOCUS AU119585 835 bp mRNA linear EST 19-OCT-2000

DEFINITION AU119585 HEMBA1 Homo sapiens cDNA clone HEMBA1006161 5', mRNA

ACCESSION AU119585

VERSION AU119585.1 GI:10934820

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and

Isogai, T.

HRI human cDNA project

Unpublished (2000)

Contact: Takao Isogai

Genomics Laboratory

Helix Research Institute

1532-3 Yana, Kisarazu, Chiba 292-0812, Japan

Tel: 81-438-52-3951

Fax: 81-438-52-3952

Email: genomics@hri.co.jp

HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix Research Institute; cDNA library construction; Department of

Virology, Institute of Medical Science, University of Tokyo, and

Helix Research Institute.

Location/Qualifiers

1.835

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="HEMBA1006161"

/clone.lib="HEMBA1"

/tissue.type="whole embryo, mainly head"

/dev.stage="embryo, 10 weeks"

/note="vector:pMEB18FJ3"

BASE COUNT 179 a 239 c 259 g 155 t 3 others

ORIGIN

alignment_scores:

Quality: 1332.50

Ratio: 5.028

Percent Similarity: 97.786

Percent Identity: 96.679

alignment_block:

US-09-653-961-2 x AU119585 ..

Align seg 1/1 to: AU119585 from: 1 to: 835

```

1  MetGlyLeuProArgLeuValCysAlaPheLeuLeuAlaAlaCysCysCys 17
19  ATGGGGCTTCCTCCAGGCTGGTCTGCGCTTCTGCTGCGCGCTGCTGCTG 68
17  sCysProArgValAlaGlyValProGlyGluAlaGluGlnProAlaProG 34
69  CTGTCTCGCGCTGCGGGGTGTGCTCCCGAGAGGCTGAGACAGCTGCGCTG 118
34  LuLeuValGluValGluValGlySerThrAlaLeuLeuLysCysGlyLeu 50
119  AGCTGGTGGAGGTGGAAAGGGGACAGACAGCCCTTGAAGTGCAGCTC 168
51  SerGlnSerGlnGlyAsnLeuSerHisValAspTrpPheSerValHisLys 67
169  TCCAGAGTCCCAAGGCAACCTCAGCCATGTGAGCTGGTTTCTGTCACAA 218
67  sGluLysArgThrLeuIlePheArgValArgGlnGlyGlnGlyGlnSerG 84
219  GGAGAAAGCGGAGGCTCATCTTCCGTGTCGCCAGGCGGAGGCGGAGGCG 268

```

84 lnpProGlyGluTyrGluGlnArgLeuSerLeuGlnAspArgGlyAlaThr 100
|||||
269 AACCTGGGAGTAGACGACGCGCTCAGCTCCAGGACGAGGCGCTACT 318
|||||
101 LeuAlaLeuThrGlnValThrProGlnAspGluArgIlePheLeuGly 117
|||||
319 CTGGCCCTGACTCAAGTACACCCCAAGACGAGCATCTCTGTGCCA 368
|||||
117 nGlyAsnArgProArgSerGlnGluTyrArgIleGlnLeuArgValTyrL 134
|||||
369 GGGCAAGCCCTCGTCCAGAGTACCGACGACGCTCCGCTCTACA 418
|||||
134 yAlaAlaProGluGluProAsnIleGlnValAsnProLeuGlyIleProVal 150
|||||
419 AACCTCGGAGGAGGCAAAATCCAGGTCAACCCCTGGGCACTCCCTGG 468
|||||
151 AsnSerLysGluProGluGluValAlaThrCysValGlyArgAsnGlyTyr 167
|||||
469 AACAGTAAGAGAGCTGAGGAGGCTGCTACTGTGTAGGAGGAGAACGGGTA 518
|||||
167 rProIleProGlnValIleTyrTyrLysAsnGlyArgProLeuLysGluG 184
|||||
519 CCCATTCTCAAGTATCTGTGACAGAAATGCGCGCTCTGAAAGGAGG 568
|||||
184 lAlAsnArgValHisIleGlnSerSerGlnThrValGluSerSerGly 200
|||||
569 AGAAGAACCGGGTCCACATTGAGTGTCCAGACTGTGAGTGCAGTGT 618
|||||
201 LeuTyrThrLeuGlnSerIleLeuLysAlaGlnLeuValLysGluAsp 217
|||||
619 TTGTACACCTTGCGAGTATTTCTGAAGGACACAGCTGTTAAAGAACAA 668
|||||
217 sAspAlaGlnPheTyrCysGluLeuAsnTyrArgLeuProSerGlyAsnH 234
|||||
669 AGATGCCAGTTTACTGTGAGCTCACTACCGGCTGCCANTGGGAAAC 718
|||||
234 isMetLysGluSerArgLeuValThrValProValPheTyrProThrG 250
|||||
719 ACATGAAGGAATTCAGGGAGTCAACGC...CTGTTCTACCCGACAG 765
|||||
250 uLysValThrPLeuGluValGluProValGlyMetLeuLysGluGlyAsp 267
|||||
766 AAAAGTGCTGCTGGAATG.GAACCGTGGGAATGCTGAAGGAAGGAC 814
|||||
267 rGValGluIle 270
|||||
815 GCGTGGAATA 825
|||||

seq_name: gp_est1:AU119127
seq_documentation_block: 743 bp mRNA linear EST 19-OCT-2000
LOCUS AU119127 HEMBA1 Homo sapiens cDNA clone HEMBA1005093 5', mRNA
DEFINITION AU119127 HEMBA1 Homo sapiens cDNA clone HEMBA1005093 5', mRNA
sequence.
ACCESSION AU119127 GI:10934362
VERSION AU119127
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 743)
Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3851
Fax: 81-438-52-3952

Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
Research Institute; cDNA library construction; Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
Location/Qualifiers
1. 743
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEMBA1005093"
/clone_lib="HEMBA1"
/tissue_type="whole embryo, mainly head"
/dev_stage="embryo, 10 weeks"
/note="vector: pME18SFL3"

BASE COUNT 185 a 189 c 238 g 128 t 3 others
ORIGIN

alignment_scores:
Quality: 1244.00 Length: 245
Ratio: 5.205 Gaps: 1
Percent Similarity: 97.551 Percent Identity: 97.551

alignment_block:
us-09-653-961-2 ..
Align seg 1/1 to: AU119127 from: 1 to: 743

216 AsplysAspAlaGlnPheTyrCysGluLeuAsnTyrArgLeuProSerG 232
|||||
3 GACAAAGTGGCCAGATTACTGTGAGCTCAACACCGGCTGCCCAATGG 52
|||||
232 yAsnHisMetLysGlnSerArgGluValThrValProValPheTyrPro 249
|||||
53 GACACCATGAAAGAGATCCAGGAAGTACCGCTGCTTTTCTACCCGA 102
|||||
249 hGluLysValThrPLeuGluValGluProValGlyMetLeuLysGluGly 265
|||||
103 CAGAAAGAGTGGCTGAGTGAAGTGGAGCCCGTGGGAATGCTGAAGAGAG 152
|||||
266 AspArgValGluIleArgCysLeuAlaAspGlyAsnProProProHisPh 282
|||||
153 GACCCGCGGAATTCAGGTGTTGGCTGATGCAACCTCCACCACTT 202
|||||
282 eSerIleSerLysGlnAsnProSerThrArgGluAlaGluGluGluThr 299
|||||
203 CAGCATCAAGCAAGCAAGACCCAGACAGGAGGAGGAGGAAGAACAA 252
|||||
299 hAsnAspAsnGlyValLeuValLeuGluProAlaArgLysGluHisSer 315
|||||
253 CCAACGACAAACGGGGTCTGCTGCTGAGCTGCCCGGAAAGAACACAGT 302
|||||
316 GlyArgTyrGluCysGlnGlyLeuAspLeuAspThrMetIleSerLeu 332
|||||
303 GGGCGCTATGAATGTCAGGGCCCTGACCTTGACACCAATATATGCTGCT 352
|||||
332 uSerGluProGlnGluLeuLeuValAsnTyrValSerAspValArgValS 349
|||||
353 GAGTAACACACAGAGAACTACTGTAATATGCTGCTGCTGCTGCTGCTG 402
|||||
349 ePrAlaAlaAlaProGluArgGlnGluGlySerSerLeuThrCys 365
|||||
403 GTCGCGAGCCCTGAGAGACAGAAAGCAGACACCTCACCTGACCTGCT 452
|||||
366 GluAlaGluSerSerGlnAspLeuGluPheGlnThrPLeuArgGluGlu 382
|||||
453 GAGGAGGAGAGTACGACGACCTCGAGTTCAGAGGCTGAGAGAGAGAC 502
|||||
382 rGlyGlnValLeuGluArgGlyProValLeuGlnLeuHisAspLeuYSA 399
|||||
503 AGCCAGAGTGTGGAAGAGGGGCGCTGCTCAAGTGTGATGATCACTGAAC 552
|||||
399 rGluAlaGlyGlyGlyTyrArgCysValAlaSerValProSerIlePro 415
|||||

```
|||||
553 GGGAGGCGAGGCGGCTATGCTGCGTGTGCGCCAGCATACCC 602
416 G1yleuAenAqTThrGlnLeuValAsnValAlaIlePheGlyProPOTR 432
603 GGCCTGAACCCACACACAGCTGTCACGTGCCCTTTTGGCCCCCTTG 652
432 pMetAlaPheLysGluArgLysValTTrpValLysGluAsnMetValIleuA 449
653 GATGGCATTTCAAGAGAGAGAGAGGTGGTGAAGANATATGNGTTGA 702
449 snIeuSerCysGluAlaIser.GlyHisProArg 459
703 ATCTGCTTGTGGAAGCCTTAAGGCGACCCCGG 735

seq_name: gb_est1:AU126003

seq_documentation_block:
LOCUS AU126003 735 bp mRNA linear EST 23-OCT-2000
DEFINITION AU126003 NT2RM4 Homo sapiens cDNA clone NT2RM4002547 5', mRNA
sequence.
ACCESSION AU126003
VERSION AU126003.1 GI:10950719
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 735)
Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Salto,K., Yamamoto,J.,
Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
Isogai,T.
HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
Salto,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki
,Y., Sugano,S., Isogai,T.)
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomis@hri.co.jp
HRI human cDNA project, 5'- & 3'-end one pass sequencing: Helix
Research Institute: cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source
1..735
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RM4002547"
/clone_11b="NT2RM4"
/cell_type="teratocarcinoma"
/note="vector: pME18SFL3; mRNA from uninduced NT2 neuronal
precursor cells"
BASE COUNT 192 a 191 c 220 g 129 t 3 others
ORIGIN

alignment_scores:
Quality: 1238.00 Length: 244
Ratio: 5.116 Gaps: 0
Percent Similarity: 99.180 Percent Identity: 99.180

alignment_block:
US-09-653-961-2 x AU126003 ..

Align seq 1/1 to: AU126003 from: 1 to: 735
131 ArgValTyrLysAlaProGluGluProGlnIleGlnValAsnProIleuGI 147
|||||
```

```
3 CGCGTCTACAAAGCTCCGGAGAGCCAAACATCCAGGTCAACCCCTGGG 52
147 YIleProValAsnSerLysGluProGluGluValAlaIleThrCysValGlyA 164
53 CATCCCTGTGAACAGTAAGAGAGCTGAGAGGTGCTACTGCTGTAGGSA 102
164 rGaSnGlyTyrProIleProGlnValIleTTrpTyrLysAsnGlyArgPro 180
103 GGAAGCGGTACCCCATTTCTCAAGTCAATCTGTGTACAGAAATGGCGGCT 152
181 LeuLysGluGluLysAsnArgValHisIleGlnSerSerGlnThrValGI 197
153 CTGAAGAGAGAGAAAGAACCGGGTCCACATTCAGTGTGCCACAGACTGTGA 202
197 userSerGlyLeuTyrThrLeuGlnSerIleLeuLysAlaGlnIleuValL 214
203 GTTCAGTGTGTTGTACACCTTGACAGATTTCTGAAGGACGACGTGGTTA 252
214 ySgLuAspLysAspAlaGlnPheTyrCysGluLeuAsnTyrArgLeuPro 230
253 AAGAAGACAAAGATGCCAGTTTACTGTGAGCTCAACTACCGGCTGCC 302
231 SerGlyAsnHisMetLysGluSerArgGluValIleThrValProValPheTyr 247
303 AGTGGAGAACCATGTAAGAGAGTCCAGGAAATCACCGCTCCCTGTTTCTTA 352
247 rProThrGluLysValTTrpLeuGluValAlaGluProValGlyMetLeuLysG 264
353 CCGCAGACAAAAGATGGCTGCTGGAAGTGAAGCCCGTGGAAATGCTGAAGG 402
264 LuGlyAspArgValGluIleArgCysLeuAlaAspGlyAsnProProPro 280
403 AAGGGGACCGCGTGAATACAGTGTGCTGATGAGCAACCTCACCA 452
281 HisPheSerIleSerLysGlnAsnProSerThrArgGluAlaGluGlu 297
453 CACTTGACGATCAGCAAGCAACCCACACGAGGAGGAGGAGGAAGA 502
297 uThrThrAsnAspAsnGlyValLeuValLeuGluProAlaArgLysGluH 314
503 GACAAACCAACGAAACGGGCTCTGCTGTGAGGCTGCCGGAAGAAC 552
314 iSerGlyArgTyrGluCysGlnGlyLeuAspLeuAspThrMetIleSer 330
553 ACAGTGGGCTATGATGATGTCAGGGCTGGAATGAGACCATGATATCG 602
331 LeuLeuSerGluProGlnGluLeuValAsnTyrValSerAspValAr 347
603 CTGCTGATGAACCAAGGAACTACTGGTGAATGTGTGTGACGCTCG 652
347 gValSerProAlaAlaProGluArgGlnGluGlySerSerLeuThrLeuT 364
653 AGTGAATCCCGCAGCCCTGAGAGACAGAGAGGAGGAGGAGCCG. ACCCTGA 701
364 hnCysGluAlaGluSerSerGlnAspLeuGlu 374
702 CCTGTGANCANANAGTACCCAGACCTCGAG 733

seq_name: gb_est1:AL550956

seq_documentation_block:
LOCUS AL550956 923 bp mRNA linear EST 16-FEB-2001
DEFINITION AL550956 LTT_NFL006_Pu2 Homo sapiens cDNA clone CS0D1065YJ10 5
prime, mRNA sequence.
ACCESSION AL550956
VERSION AL550956.1 GI:12888440
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 923)
Ll,W.B., Gruber,C., Jessee,J. and Polayes,D.
```

TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source 1..923
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSOD1065XJ10"
/clone_lib="LIT_NFL006_PL2"
/tissue_type="Placenta"
/note="Vector: pCMVSPORT 6; Site: 1. NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 202 a 212 c 268 g 192 t 49 others
ORIGIN

alignment_scores:

Quality: 1221.00 Length: 304
Ratio: 4.556 Gaps: 1
Percent Similarity: 88.158 Percent Identity: 84.868

alignment_block:
US-09-653-961-2 x AL550956 ..

Align seg 1/1 to: AL550956 from: 1 to: 923

```
1 MetGlyLeuProArgLeuValGlyAlaPheLeuAlaAlaCysCysCys 17
|||||
1 ATGGGGCTTCCAGGCTGGTCTGCGCTTCTTCTGCGCGCGCTGCTG 50
17 scysProArgValAlaGlyValProGlyGluAlaGluInProAlaProG 34
|||||
51 CTGTCTCTGCGCTGCGGGGTGTGCGCGAGAGGCTGAGACCCCTGCTG 100
34 IuLeuValGluValGluValGlySerThrAlaLeuLeuLysCysGlyLeu 50
|||||
101 AGCTGCTGAGGAGTGGAGAGTGGGACACACACCTTCTGAGAGTGGGCTC 150
51 SerGlnSerGlnGlnValAsnLeuSerHisValAspTrpPheSerValHis 67
|||||
151 TCCCACTCCCAAGGCAACCTCAGCCATGTCAGTCTGTTCTGTCACAA 200
67 sGlyGlyArgThrLeuIlePheArgValArgGlnGlnGlnGlnSerG 84
|||||
201 GGGAGAGCGGAGCGCTCATCTCTCGTGTGCGCAGGGCGGCGCAAGCG 250
84 IuProGlyGluTrpGlnGlnArgLeuSerLeuGlnAspArgGlyAlaThr 100
|||||
251 AACCTGGGAGTAGACGACGCGCTCAGCTCCAGAGANANAGGCGCTACT 300
101 LeuAlaLeuThrGlnValThrProGlnAspGluArgIlePheLeuGly 117
|||||
301 CTGGCGCTTCAATCAAGTACCCCAAGATGAGCGATATTTTTRTGCCA 350
117 nGlyGlyArgProArgSerGlnGluTrpArgIleGlnLeuArgValTyr 134
|||||
351 GGGCAAGCGGCTCTCGGTTCAGGAKTAYGCAATYAGTYCGGTYTAYA 400
134 yAlaAlaProGluGluProAsnIleGlnValAsnProLeuGlyIleProVal 150
|||||
401 AAKYTCCGAGAGATYA. AAYATTYAGTTAACYCCCTGGGCAATYTTGTT 449
```

```
151 AsnSerLysGluProGluGluValAlaThrCysValGlyArgAsnGlyTyr 167
|||||
450 AATATKTAAGAGAGCTTGAAGAGTGGCTACTGTTGTAGGAGGAATGCTGA 499
167 rProIleProGluValIleTrpTyrLysAsnGlyArgProLeuLysGluG 184
|||||
500 TCCCATYTYTYAMTTTATTTTGTACAGAAATGTGTGTTCTGTGAMGATG 549
184 IuLysAsnArgValHisIleGlnSerSerGlnThrValGluSerSerGly 200
|||||
550 AGAGMWTGTTGGTCTATATTATTTAGTGTCCAGACTGTGAGTGAATGTT 599
201 LeuTrpThrLeuGlnSerIleLeuLysAlaGlnLeuValLysGluAsp 217
|||||
600 TTGATACCTTGTGAGATATTCTGAAGCAGACAGTGTTAAGAGACAA 649
217 sAspAlaGlnPheTyrCysGluLeuAsnTyrArgLeuProSerGlyAsn 234
|||||
650 AGATGCCCACTTTTACTGTAGCTCACTACCGGCTGCCAGTGGGAA. C 698
234 IsmLysGluSerArgGluValThrValProValPheTyrProThrGlu 250
|||||
699 ACATGAAGAGAGTCCAGGAGATCACGTCCTGTTTCTACCGAATAGAA 748
251 LysValTrpLeu. GluValGluProValGlyMetLeuLysGluGlyAsp 267
|||||
749 AAGTGTGTGTGGAAKTGKAGYCCGTGGGAATGCTTAAGAGAGGAGC 797
267 rGValGluIleArgCysLeuAlaAspGlyLysnProProHisPheSer 283
|||||
798 C. GTGGAATCAGGTTGTTGCTGATGCAACCTCCACACACTTCAGC 847
284 IleSerLysGlnAsnProSerThrArgGluAlaGluGluInLthrTrpAs 300
|||||
848 ATCAGCAGCAGAGAACCCAGCACCGAGGAGGAGAGAGAAACAAAC 897
300 nAspAsnGly 303
898 GACACAGCGG 907

seq_name: gb_est1:AU130835
seq_documentation_block:
LOCUS AU130835 735 bp mRNA linear EST 24-OCT-2000
DEFINITION AU130835 NT2RP3 Homo sapiens cDNA clone NT2RP3001526 5', mRNA
ACCESSION AU130835
VERSION AU130835.1 GI:10991189
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 735)
Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Ishigaki,T.
Isogai,T.
HRI human cDNA project
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genom@hri.co.jp
HRI human cDNA project; 5'- & 3'-end pass sequencing; Helix
Research Institute; cDNA library construction; Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
location/Qualifiers
1..735
/organism="Homo sapiens"
```

```

/db_xref="taxon:9606"
/clone="NT2RP3001526"
/clone_id="NT2RP3"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/Note="Vector: pME18SF13; mRNA from NT2 neuronal precursor
cells after 2-weeks retinoic acid (RA) induction"
BASE COUNT      155 a      221 c      221 g      135 t      3 others
ORIGIN

```

```

alignment_scores:
    Quality: 1193.00      Length: 232
    Ratio: 5.210          Gaps: 0
Percent Similarity: 98.707 Percent Identity: 98.276

```

alignment_block:

US-09-653-961-2 x AU130835 ..

Align seg 1/1 to: AU130835 from: 1 to: 735

```

1 MetGlyLeuProArgLeuValGlyAlaPheLeuAlaAlaCysCysC 17
|||||
37 ATGGGGCTTCCAGAGCTGCTGCTGCTTCTGCTGCTGCTGCTG 86
|||||
17 SCYPRoArgValAlaGlyValProGlyGluAlaGluGlnProAlaProG 34
|||||
87 CTGCTCCTCGGCTCGGCTGCTCGGAGAGGCTGACACCTCGGCTG 136
|||||
34 LuLeuValGluValGluValGlySerThrAlaLeuLeuLeuGlyLeu 50
|||||
137 AGCTGGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 186
|||||
51 SerGlnSerGlnGlyAsnLeuSerHisValAspTrpPheSerValHis 67
|||||
187 TCCAGTCCCAAGGCAACCTCAGCCATGTCGACTGGTTTCTGTCACAA 236
|||||
67 sGluysArgThrLeuLeuPheArgValArgGlnGlnGlnSerG 84
|||||
237 GGAGAACGGGAGCTCATCTTCGCTGCGCCAGAGGCGAGGCGAGGCG 286
|||||
84 LuProGlyGlyTrpGluGlnArgLeuSerLeuGlnAspArgGlyAlaThr 100
|||||
287 AACCTGGGAGTACGACGAGCTGCTCAGCTCCAGGACAGAGGCTACT 336
|||||
101 LeuAlaLeuThrGlnValThrProGlnAspGluArgTLeuPheLeuGly 117
|||||
337 CTGGCCCTGACTCAAGTCACCCCAAGAGAGGCGCATCTTCTTGTCGA 386
|||||
117 nglyysArgProArgSerGlnGlyTrpArgTLeuGlnLeuArgValTrp 134
|||||
387 GGGCAACGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 436
|||||
134 ysaAlaProGluGluProAsnTLeuValAsnProLeuGlyTLeuProVal 150
|||||
437 AAGCTCGGAGAGGCAAAACATCCAGGTCAACCCCTGGGCAATCCCTG 486
|||||
151 AsnSerLeuGluProGluGluValAlaThrCysValGlyArgAsnGlyTrp 167
|||||
487 AACAGTAAAGAGGCTGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 536
|||||
167 rProAlaProGlnValTLeuTrpTrpTrpTrpTrpTrpTrpTrpTrp 184
|||||
537 CCCCATTCCTCAAGTCAATGCTGCTCAAGAAATGGCCGCTCTGAAAGAG 586
|||||
184 LuysAsnArgValHisTLeuGlnSerGlnThrValGlnSerSerGly 200
|||||
587 AGAAGAACCGGCTCAGATTCAGTCCAGACTGTGAGTCAAGTGGG 636
|||||
201 LeuTrpThrLeuGlnSerTLeuLeuValAlaGlnLeuValGlyGluAsp 217
|||||
637 TTGTACACCTTGCAGAGTATTCTGAAAGGCACTGGTTTAAGAGACAA 686

```

```

217 sasAlaGlnPheTrpCysGluLeuAsnTrpArgLeuProSerGly 232
|||||
687 AGATGCCCAATTTTACTGAGCTCACTACCGGTTGCANTGGGA 732

```

seq_name: gb_est1:AL546898

seq_documentation_block:

LOCUS AL546898 773 bp mRNA linear EST 16-FEB-2001
DEFINITION AL546898 LTI_NFL006_P12 Homo sapiens cDNA clone CS0D1026YB05 5
prime, mRNA sequence.

ACCESSION AL546898 GI:12880463

VERSION AL546898.1

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 773)
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

AUTHORS 'L.W.B., Gruber,C., Jessee,V. and Polyes,D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..773
Location/Qualifiers

```

/db_xref="taxon:9606"
/clone="CS0D1026YB05"
/clone_id="LTI_NFL006_P12"
/lssue_type="placenta"
/Note="Vector: PCWSPORT 6; Site: 1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the PCWSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT      166 a      227 c      205 g      163 t      12 others
ORIGIN

```

```

alignment_scores:
    Quality: 1159.50      Length: 255
    Ratio: 4.872          Gaps: 2
Percent Similarity: 93.333 Percent Identity: 90.588

```

alignment_block:

US-09-653-961-2 x AL546898 ..

Align seg 1/1 to: AL546898 from: 1 to: 773

```

1 MetGlyLeuProArgLeuValGlyAlaPheLeuAlaAlaCysCysC 17
|||||
11 ATGGKCTTCCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 60
|||||
17 SCYPRoArgValAlaGlyValProGlyGluAlaGluGlnProAlaProG 34
|||||
61 CTGCTCCTCGGCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 110
|||||
34 LuLeuValGluValGluValGlySerThrAlaLeuLeuLeuGlyLeu 50
|||||
111 ATCTGTTGAGAGTGGAAATTTGGGAGAGAGCCCTTTAAGTGGGCTC 160
|||||
51 SerGlnSerGlnGlyAsnLeuSerHisValAspTrpPheSerValHis 67
|||||
161 TCCAGTCCCAAGGCAACCTCAGCCATGCTGCTGCTGCTGCTGCTGCTG 210
|||||
67 sGluysArgThrLeuLeuPheArgValArgGlnGlnGlnSerG 84

```



```
|||||
421 AACCTCCGAGAGAGCAACATCCAGSTCAACCCCTGGGCATCCCTGTG 470
151 AsnSerLysGluProGluGluValAlaThrCysValGlyArg.AsnGlyT 167
471 AACAGTAAGAGAGCCTGAGAGGTCGTACCTGTGTAGGAGGGAACGGGT 520
167 YrProIleProGluValIleTyrTyrLysAsnGlyArgProLeuLysGlu 183
521 ACCCATTTCTCTCAAGTCATCTGGTACAAAGAATGGCCGCCCTGTGAAGAG 570
184 GluLysAsnArgValHisIleGlnSerSerGlnThrValGlu.SerSerG 200
571 GAGAAAGAACCGGCTCCACATTCAGTCCGCAAACTGTGAGATNCGAGTG 620
200 LysLeuTyrThrLeuGlnSerIleLeuLysAlaGlnLeuValLysGluAsp 216
621 GTTGTGACACCTTGACAGATATTCTGAAGGCACAGCTGTTAAAGAAAGA 670
217 LysAspAlaGlnPheTyrCysGluLeuAsnTyrArg 228
671 CAAGTCCCAAGTTTACTGTGTGAGCTCACTACCGG 706
```